

Money turmoil poses awkward questions

The world's foreign exchanges have been like bear-pits in the past few weeks, but the causes of the turmoil have long roots, in the US budget deficit and in most of Europe's tradition of tolerating inflation.

HERE is an intellectual problem with practical as well as intellectual interest. Germany is in money trouble because of the Bonn government's failure to estimate the full social and money costs of reunification and because of the threat of inflation, signalled by the growth of the domestic money supply last month. Yet for the past three weeks, the Deutschmark has been riding high on the world's foreign exchanges — embarrassingly so. In the United States, on the other hand, inflation is lower than for decades and still falling, while the Federal Reserve has moved vigorously to stimulate economic growth by reducing interest rates to historically low levels. Yet the dollar has fallen precipitously in value against the Deutschmark (by roughly 20 per cent since the beginning of this year). A dollar will now buy much more in New York than will DM1.47 in Frankfurt; with even sterling stronger against the dollar (but weaker against the Deutschmark), the British Customs and Excise have thought it prudent to warn British travellers to the United States that they will have to pay tax on whatever bargains they find in North America. Why the difference?

The simple explanation is that German interest rates are now six or seven percentage points above those obtainable in the United States. Not only foreign exchange dealers, but companies and private persons seeking to make the best of their bank deposits, switch from accounts denominated in dollars to accounts denominated in Deutschmarks for the sake of the extra interest their money will earn. But that is only part of the story, and in particular does not explain why the flight from the dollar has now gone so far that it would probably be possible for German traders to satisfy the national demand for *bratwurst* and other such goods by buying in Chicago with cheap dollars and then paying the transport costs to Hamburg. (There would, of course, be European Community tariffs to pay, which is another tale.)

The more durable cause of the weakness of the dollar in the past few weeks is different, and is to be found in Washington. Estimating the annual US federal deficit is one of the few growth industries in the United States. Whether the final figure for this financial year (which ends this month) will be nearer to \$400 billion than \$500 billion remains conjecture. But it is now clear that countries such as Germany and Japan, which earn contentiously respectable trading surpluses, have other things to do with the proceeds of their surpluses than to return them to the United States to be used, directly or otherwise, to bridge the federal government's budget gap. The Federal Reserve's drive towards low

interest rates can only have confirmed the resolve of foreigners not to help out the US administration. It is remarkable that none of this can be gleaned from the election campaign now under way.

The strength of the Deutschmark, relative to other European currencies as well as the US dollar, must be differently explained. Indeed, in the best sense of the phrase, it is a confidence trick. Over the years, the central bank (called the Bundesbank) has acquired such a fearsome reputation for monetary probity, confirmed by the increase of interest rates last month, that the investment community believes that eight per cent interest on Deutschmark deposits is a safer risk than, say, ten per cent in Britain or France or even seventeen per cent in Italian lire. In the past few weeks, the money changers have been betting on a realignment of European currencies — on a stronger Deutschmark and weaker currencies elsewhere, which would bring them capital profits as well as higher earnings.

How safe is that bet? If the street violence in Rostock over recent days had happened elsewhere, the local currency would no doubt have tumbled. But the exchange markets might also have been given pause by last week's decision by a US chipmaker to close its plant at Braunschweig on the grounds that German labour costs and practices make it unprofitable to operate expensive capital equipment there. (That, in the long run, will be the whole of Europe's problem.) More immediately, there is the chance that France may say *NON* at the referendum on the Maastricht Treaty less than three weeks from now; by illogically casting doubt on the viability of the whole European enterprise, that could make all European currencies less valuable against the dollar. But if the project for European monetary union has to be abandoned or postponed, it is difficult to see how revaluation of the Deutschmark could be avoided. That is what the European bets are about.

FEMA's testing time

The US federal agency with responsibility for national emergencies has flunked again.

If emergency management, like charity, begins at home, then the victims of the hurricanes in Florida this season are unlikely to get speedy succour from the Federal Emergency Management Agency (FEMA). That body has once again



failed to rise to one of the occasions for which it was formed by its dithering over the damage caused in Florida by last week's hurricane, but instead added another chapter to the record of its reputation of unhelpfulness begun shortly after its creation in 1979.

FEMA, although conspicuous, is not alone among the government entities that appear to have been laggard in dealing with the residents of South Florida battered by Hurricane Andrew last week. The governor of Florida conceded that he had underestimated the severity of the hurricane, and White House spokesmen in Washington said that they began moving swiftly once they were told by state officials that relief work was too much for the Florida National Guard or Army Reservists. But that happened four days after the storm crunched its way across Florida and left some 200,000 people homeless in a welter of destruction and wreckage one witness called "worse than Kuwait City" after the Persian Gulf War.

It is not that nobody knew in advance about Hurricane Andrew and its viciousness, but it seemed as if not many people were telling others. The word did not float up to levels of officialdom where its significance could be appreciated, and, by the time it did, uncounted tens of thousands of people had gone without reliable water and electricity in 98-degree temperatures (Fahrenheit) for four days. By then, the people were mightily disgruntled. Mere promissory words from government offices would not suffice. What made the government's neglect grate more was that truckloads of food, ice, clothing, tools, and other necessities began pouring in from church groups, the American Red Cross, and other organizations as far away as Ohio.

FEMA held a press conference to assure reporters that it would fulfil its mission of producing a "coordinated" effort from among the 27 federal agencies whose help it can enlist in disaster relief programmes. The director of FEMA did not attend the press conference, but shortly afterwards the White House announced that the task of coordinating the relief work would be handled by the Secretary of Transportation; it was not immediately clear what the FEMA director would do. Longtime critics of FEMA doubted that the entire agency would do much, because historically it never has.

FEMA was formed during the Carter administration to serve as evaluator and disbursing officer of relief funds and to plan civil emergency preparedness for nuclear attack, among other things. The budget request for the impending fiscal year is at the billion-dollar level, but there is legislative work afoot to cut it by a couple of million. There is much conviction in Congress that FEMA has been ingloriously tardy and undersupplied in such disasters as Hurricane Hugo and the San Francisco earthquake, both in 1989, and such other events as the Los Angeles riots of 1992.

Those familiar with FEMA's shortcomings say they doubt that the sheer size of the Andrew devastation will be met by civil "coordination" efforts led by FEMA. It would take an army to meet the task, they say. Why not the US Army, which worked out the logistics of Desert Storm and the feeding of 500,000 troops in an exercise in central Florida? □

AIDS and virginity

The New York City Board of Education is in the dark ages, demanding that AIDS education emphasize abstinence.

It is difficult to imagine that, in what is supposed to be one of the most sophisticated cities in the world, serious people are still unwilling to recognize, first, that teenagers have sexual intercourse and, second, that the relentless spread of AIDS is aided and abetted by unprotected sexual contact. But that is exactly the case in New York City where, to the consternation of AIDS educators and the city Health Commissioner, the Board of Education has just voted for a policy requiring that anyone who enters the school system to teach about AIDS must sign a pledge promising to "stress that abstinence is the most appropriate premarital protection against AIDS". Further, the AIDS educators must pledge in writing that they will "devote substantially more time and attention to abstinence than to other methods of prevention".

For all practical purposes, this official break with reality and common sense means that AIDS education will have to be offered outside the school system, which some administrators opposed to AIDS education say is just fine with them. Displaying unrepentant homophobia, one is quoted as saying of the outside groups offering AIDS education that "most of these people going in there are the gay and lesbian community".

The fact that Health Commissioner Margaret A. Hamburg, backed by New York Mayor David N. Dinkins, is opposed to the moralistic pledge is apparently of little concern to the education board. Dr Hamburg is lending moral support to people who refuse to sign the pledge, rightly saying that AIDS education should be based on science and medicine. Unfortunately, she has no authority in the matter.

Meanwhile, AIDS education has also taken a knock in Washington, where the US National Commission on AIDS, whose members include basketball star Magic Johnson, has had to lay off half of its 20-member staff and cancel planned meetings because of budget restraints. The commission, counting on the use of \$750,000 from last year's budget, has just been told by the General Accounting Office that the money cannot be used in this fiscal year.

The national commission has been unrealistically harsh in its criticism of the Bush administration's response to the AIDS crisis, calling it "tragically insufficient". In fact, federal spending on AIDS research and drug testing is greater than for any other single disease. But the commission is entirely right in its criticism of the administration's benighted position on AIDS education. Though not as extreme as the New York City Board of Education, the administration, with its politically driven commitment to "family values", also refuses to accept that people practising "safe sex" (that is, the use of a condom) are less likely to get AIDS than those who have unprotected sex with several partners (the argument that condoms are not 100 per cent effective notwithstanding). The tragedy in this clash among so-called adults about what to teach on AIDS is that it is young people who will suffer. □

Paralysed by politics, EPA delays spending any money on EMF research

Washington. Research on the possible health risks of electric power lines and other sources of electromagnetic fields (EMFs) is so controversial that at least one US science agency has refused even to spend the research money it has been given. Officials at the Environmental Protection Agency (EPA) say that they lack a coordinated research programme, but other evidence suggests that the agency may be simply avoiding a political hot potato.

Although the US Department of Energy (DOE) has also delayed EMF funding out of fear of lending credence to those who say the problem is grave, critics of the government's research policies have focused on EPA, which received \$1.9 million from Congress for the fiscal year ending this month. That is a tiny sum compared to EMF research budgets elsewhere in the government and in industry, but not a penny of it has yet been spent.

Why is an agency with both an in-house research effort and a relatively large grants programme unable to spend less than \$2 million? Many people want to know.

The unspent money has drawn the attention of Representative George Brown (Democrat, California), chairman of the House science committee, and Representative James Scheuer (Democrat, New York), chairman of its subcommittee on the environment. They plan to send a letter to EPA asking about the funds when Congress returns next week from a month-long recess. The money has also attracted the notice of Senator Joseph Lieberman (Democrat, Connecticut), who told EPA last month that he might try to shift the money to the National Institute for Occupational Safety and Health for studies of cancer among police officers who use radar detectors. And a newsletter, *Microwave News*, entitled an editorial "Fear and Cowardice at EPA".

EPA officials say they have not spent the money because they do not want to waste it on marginal research. David Kleffman, deputy director of health research in the EPA Office of Research and Development, says that EPA has been reluctant to act without discussing the issue with a government-wide committee that includes officials from DOE, the White House and the Department of Health and Human Services. "The whole process, just because of the number of players, has turned out to be very cumbersome", he says.

But most observers believe that there is more to it than that. EPA started an EMF research programme in 1983 that was cancelled from above in 1986. At its peak, EPA had 27 full-time researchers and a budget of

\$3.5 million, almost as much as DOE, which has traditionally led the federal effort. Today, only a few EPA researchers are investigating EMF effects, working in such related fields as radiofrequency radiation.

Part of the reason for EPA's research troubles is that powerful officials in the administration and the scientific community doubt that EMF radiation poses a health risk. These sceptics include D. Allan Bromley, presidential science adviser, and Robert Adair, a prominent physicist and Bromley's former colleague at Yale Univer-



sity, who fear that the public will interpret federal funding as proof that a problem exists. It could also force utility companies and electronics manufacturers to spend millions of dollars to relocate power lines and to shield equipment. Bromley has met several times with EPA officials on the subject, especially after the agency released a controversial 1990 report labelling EMFs as "probable" carcinogens, and he has reportedly asked the agency to be less aggressive in pursuing the issue.

Ironically, the issue of EMFs finds industry, consumer activists and environmental groups in rare agreement: they all want more government research. The industry-funded Electric Power Research Institute (EPRI) is the largest player in the field, having spent about \$50 million, with plans to spend another \$15 million in 1992. But although EPRI is considered relatively independent and has avoided most charges of industry bias, utility officials are convinced that the public will believe research dis-

counting an EMF threat only if it comes from outside the industry, preferably from government-funded independent research.

Last year, Congress forced EPA back into the field by giving it \$800,000 that it did not request; the agency passed most of it on to the Health Effects Institute (HEI) in Cambridge, Massachusetts, for preparation of a research plan for an EMF research programme. But that approach backfired when HEI, an independent research organization funded half by industry and half by government and other grants, became enmeshed in a controversy over a report on asbestos risks. (Critics accused it of packing its scientific advisory panel with researchers connected with industry.) Although the final report was well received, the incident cost the institute its president and some of its credibility.

Seeking to avoid a similar controversy, HEI is moving ahead very slowly on the EMF project. Charles Powers, acting HEI president, says that the institute has convened a new advisory panel and expects to finish the report by the end of the year, more than a year late. And EPA, which had planned to give HEI the 1992 funding to set up an EMF research programme, has not yet sought another place to spend the money.

But while EPA has had so much trouble spending less than \$2 million, DOE has managed to operate a \$5.8-million EMF research programme with minimal controversy. It has even asked for an increase to \$7.5 million in fiscal year 1993. DOE plans to withhold spending \$1.5 million of its 1992 funding (its entire increase over 1991) until it receives the recommendations next month of an interagency steering group, but it is not expected to have trouble finding takers once it moves ahead.

Why is one agency being encouraged to expand its programme while another's efforts are being discouraged? Keith Florig, an analyst with Resources for the Future, believes that DOE's effort is simply less controversial. Most of DOE's EMF programme is in-house research on cellular models, with almost no funding for epidemiology. To date, the only research that has consistently provided cause for alarm has been epidemiological in nature.

"EPA comes at things with a presumption of environmental problems", says one congressional aide. "DOE, on the other hand, has an historic tendency to find nothing wrong — and in EMF they are simply not addressing the most relevant research questions." In this administration, this aide concludes, the best EMF research appears to be that which leaves the least wake.

Christopher Anderson

Big increase for MITI budget emphasizes energy technology

Tokyo. Japan's Ministry of International Trade and Industry (MITI) last week applied for its largest budgetary increase for research in more than a decade.

The increase of 16 per cent for the fiscal year beginning 1 April 1993 would come largely from a new special account of ¥142 billion (\$1.1 billion) created by the ministry to promote development and use of technology to reduce carbon dioxide emissions and energy consumption. The ministry will use surplus funds and borrow money to finance the account in 1993, but the Japanese government must impose some new form of tax in subsequent years.

MITI's research and development budget has increased in recent years by only a few per cent a year, barely enough to keep pace with inflation. The ministry has been forced to reduce its general expenditures under a policy aimed at reducing the deficit, and its budget has grown only through separate special accounts financed by taxes on oil and electricity. The extra money in 1993 will come from a new special account.

The budget request must be screened by the Ministry of Finance and approved by the Diet before it goes into effect. But few changes are likely.

Of the ministry's overall request of ¥900.6 billion, ¥282.7 billion will go towards efforts to save energy and to protect the environment. This increase of 55 per cent is made possible with the new account. Much of the extra money will go into subsidies to encourage industry to reduce its consumption of energy, to recycle goods and to introduce substitutes for chlorofluorocarbons (CFCs).

The ministry's proposed research and

development budget of ¥301.3 billion includes a consolidation of projects related to energy and the global environment under one umbrella called the New Sunshine Project. This project, with a proposed budget of ¥55.7 billion, is the successor to the Sunshine Project (started in 1974 to develop alternative sources of energy), the Moonlight Project (initiated in 1978 to develop technology to conserve energy) and the ministry's recent projects to develop environment-friendly technology. But the research budget also includes several new initiatives, among them a proposal to spend ¥50 billion in the next seven years on technology to use some of the heat generated from cities.

The ministry plans to spend ¥300 billion during the next 27 years on an international project to develop 'clean' hydrogen energy. The idea is to develop technology to help countries rich in hydroelectric energy, such as Canada, to extract hydrogen from water by electrolysis and to sell the energy.

The special account would also support a ¥23.3-billion programme to sell environment-friendly technology developed in Japan, such as desulphurization plants, to developing countries. And the recently formed Research Institute of Innovative Technology for the Earth (RITE), which is developing substitute CFCs and ways to absorb CO₂ with help from Japanese industry, gets a boost of from ¥5–7 billion.

A long-term source of funds must be found to maintain such programmes. The three leading candidates are an environment (or CO₂) tax on energy use, a tax to support international activities including research and a tax on crude oil imports like that temporarily imposed to fund Japan's \$9-billion contribution to the Gulf War.

Apart from energy-related research, MITI would like to spend ¥3.6 billion to get its new Real World Computing Project, the follow-up to the fifth-generation computer project, into full swing next year. Its centre will be a new research institute in Tsukuba science city linked by a high-speed computer network to 20 satellite research centres in Japan and to several overseas.

The project is intended to develop new massively parallel computers, optical computers and neural network computers. In contrast to the fifth-generation computer project, which relied on the

domestic development of technology, the Tsukuba centre will lease a parallel computer from Thinking Machines Inc. of Cambridge, Massachusetts. Next month, German researchers will become full participants in the project, soon to be joined by researchers from the Korea Advanced Institute of Science and Technology in South Korea. But the United States will be involved only in one small part of the optical computing section (see *Nature* **356**, 734; 1992).

The fifth-generation computer project, which by most accounts has been a failure,

More for space and genome database

Tokyo. There are few surprises in the budget request released last week by Japan's Science and Technology Agency. Spending on space would grow by nearly 10 per cent to pay the rising costs of Japan's contribution to US Space Station Freedom and the increased development costs of Japan's next-generation H-II rocket, plagued with accidents and delays. The Human Genome Project appears to receive a large increase, but a substantial portion would go for a computer link with the Genome Data Base at Johns Hopkins University in Baltimore.

D.S.

has been extended with a shrunken budget of ¥1.4 billion. This will support 30–40 researchers, about 30 per cent of the current number. The goal is to make the fifth-generation computer compatible with standard computers with UNIX operating systems to increase its applications (*Nature* **357**, 619; 1992). The ministry has also asked for ¥1.2 billion to begin its Intelligent Manufacturing System (IMS) project, although negotiations are still under way with the United States, Europe and other partners to shape this international effort to develop the automated factories of the future.

Another significant change in next year's budget is the joining of MITI's large-scale projects to promote industrial development of new technologies with smaller next-generation industrial projects. Their budgets are combined into a new category called Industrial Scientific Technology.

MITI's search for alternative sources of energy has led it to back some long shots. The ministry's Natural Resources and Energy Agency has requested ¥300 million for technology to extract heat from cold-fusion experiments in which an electric current is passed through palladium electrodes immersed in heavy water. The agency hopes to spend ¥2.7 billion during the next four years on applications of cold-fusion technology (see *Nature* **358**, 268; 1992).

David Swinbanks

Some highlights of MITI's R&D budget

(in billion yen, ¥125 = \$1)

	1993 proposed	% change from 1992
Total R&D budget	301.3	+16.0
including such programmes as:		
Japan Key Technology Centre	29.0*	+1.7
Industrial Scientific Technology	25.9	+9.7
New Sunshine Project	55.7	+10.7
Green Aid Plan	23.3	+862.9
Real World Computing Project	3.6	+400.0
Fifth-generation Computer	1.4	-62.0
Intelligent Manufacturing System Project	1.2	+50.0
Human Frontier Science Programme	4.0**	+4.0
NEDO International Grants	0.9	+35.3
Hydrogen Energy (cold fusion)	0.3	—

*Budget shared with Ministry of Posts and Telecommunications

**Budget shared with Science and Technology Agency

BAAS embraces role of educating public

Southampton, UK. Last week's annual meeting of the British Association for the Advancement of Science (BAAS) at the University of Southampton marks a watershed in the organization's 160-year history, its transition from a broad-based scientific meeting to a public science festival. Not only is this year the first time the annual meeting has explicitly been called a festival, but the organization is taking several steps over the next 12 months to bring science even closer to the public.

One of its goals is to become more visible, with activities throughout the year rather than chiefly in connection with its annual meeting. It also wants to raise the profile of its magazine, *Science and Public Affairs*, which is published with the Royal Society. To do so, it hopes nearly to double its annual revenue, from £873,000 (US\$1.7 million) this year to £1.54 million in 1996–97.

Broadcaster Sir David Attenborough, in a presidential address entitled "Science for the public and a public for science", announced three donations that will help to change the way the BAAS operates. Sir John Cass's foundation has donated £60,000 to support a youth development officer and the Rutherford Trust £150,000 to promote the organization's activities around the country, in particular linking regional centres for interactive instruction. The largest gift is

from the Wellcome Trust; its £265,000 donation will be used to raise the BAAS's profile generally.

The head of that new effort will be Brian Gamble, now director of the popular Edinburgh Science Festival. Gamble, recalling the place that science held as a subject of drawing-room conversation 100 years ago, believes that the BAAS needs to make scientific issues more culturally respectable. Such respectability, he says, will promote the BAAS and increase donations from corporations.

Whether or not Gamble succeeds, an exchange of views of the kind that occurred during past annual meetings — such as the 1860 debate between Thomas Huxley and Soapy Sam Wilberforce, Bishop of Oxford, on the origin of species — is unlikely to return. The controversies that have arisen in recent years and at last week's meeting seem to be of greater interest to the media than to scientists.

A day-long session on a theory that man developed from an aquatic hominid ancestor took place outside the main anthropology and archaeology programme without substantial input from conventional theories of human evolution. Originally planned as a separate conference to be hosted by the university, the speakers were drafted into the BAAS meeting, at Attenborough's in-

sistence, with only 10 days' notice.

Although section recorder Malcolm Smith of the University of Durham said that the session might have been stronger if it had been part of the anthropology programme, he said that it accomplished its chief aim by focusing attention on the subject without detracting from the mainstream presentations. Attenborough said that proponents of the aquatic-ape theory may lack scientific rigour but did manage to generate a lot of interest.

With the exception of a noisy press conference, the work of British chemist Martin Fleischmann on cold fusion also went undebated after plans were abandoned to have him appear before a panel of scientists. Fleischman delivered a well-attended public lecture, during which he showed a videotaped example of what he and Stanley Pons claim is cold fusion, but he received almost no questions afterwards on the technical aspects of his work despite the presence in the audience of a number of distinguished chemists.

Few scientists see the annual meeting as an opportunity to add to their knowledge. Rather, they consider engaging the media and attracting young people into science to be far more important than the chance to present original research to their peers at an interdisciplinary meeting. **Ian Mundell**

Rubbia prods Japanese to pay more to CERN

Tsukuba, Japan. The European Laboratory for Particle Physics (CERN) is increasing pressure on Japan to pay for its growing use of CERN facilities. Carlo Rubbia, director-general of CERN, visited Japan last week to cajole Japanese scientists and government officials into contributing money to CERN, in particular the laboratory's planned Large Hadron Collider (LHC).

Rubbia's visit follows one in June by CERN officials (see *Nature* 357, 429; 1992) and is intended to counter an effort by US officials on behalf of the \$8.5-billion Superconducting Super Collider (SSC), which has similar goals to the LHC. Rubbia hopes to tap into the strong support that CERN enjoys among Japan's high-energy physics community, which is not enthusiastic about the SSC (see *Nature* 358, 266; 1992). But Rubbia will have a harder time with Japan's politicians.

Rubbia attended a joint symposium on future projects of Japan and CERN at the National Laboratory High Energy Physics (KEK) in Tsukuba science city. He won support from the 100 Japanese physicists at the symposium for increased collaboration with CERN, and in exchange offered to help Japan with its plans to build a B-meson

factory (see *Nature* 358, 266; 1992).

The only sour note in the two-day symposium was sounded by Takahiko Kondo, the organizer of Japanese participation in the SSC, who repeatedly raised questions about the technical feasibility of certain aspects of the LHC. "I am disappointed about the development of the 10T magnet [for LHC]", Kondo said in concluding remarks at the end of the symposium. "Magnets for big colliders are usually designed conservatively ... but your approach is very aggressive. I worry that you rely too much on high technology from the beginning."

But Rubbia said that it does not matter if CERN does not achieve 10T. "It's a soft figure", he says, "and nothing goes wrong if we end up with 8.5 Tesla". Rubbia then counterattacked by suggesting that the \$200,000 estimate for the cost of each of the SSC's nearly 10,000 superconducting magnets is too low.

Apart from Kondo, those in the audience favoured increased collaboration with CERN in part because they believe that CERN is offering them a much better deal. CERN is seeking 20 per cent, or about \$300 million, from non-member countries including Japan. In contrast, the United States is asking

Japan for \$1.5 billion to help build the SSC, and many Japanese physicists fear that such a payment will harm domestic projects.

After the symposium, Rubbia and his colleagues visited all the science-related ministries and agencies in Tokyo, where they undoubtedly faced a more sceptical audience. Japan's prime minister, Kiichi Miyazawa, has inched his country towards support of the SSC by agreeing to a joint United States–Japan working group on the US collider, and it is hard to imagine how Japan could afford to contribute to both machines.

But Rubbia emphasizes that CERN is not specifically seeking money for the LHC. Although almost 100 Japanese scientists regularly use CERN facilities such as the Large Electron–Positron (LEP) collider, Japan has no formal framework for collaboration and makes no financial contribution for the growing number of Japanese scientists coming to Geneva.

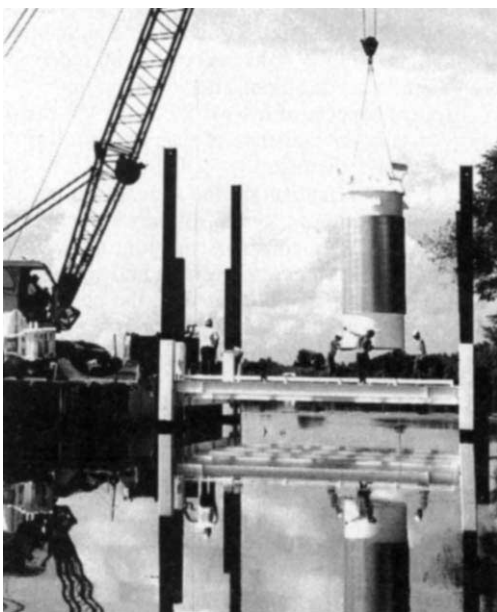
In the meantime, Japan's informal collaboration with CERN is likely to increase. "There are many things you can do without money or government approval", Rubbia said in Tsukuba, citing joint research on components for the LHC or the B-factory.

David Swinbanks

US rules hinder research on disposal of PCBs

Washington. The US Environmental Protection Agency (EPA) is reviewing its restrictions on the use of toxic chemicals known as polychlorinated biphenyls (PCBs). The new rules could make it easier and less expensive for scientists to conduct research on removing PCBs from the environment.

The current regulations are especially bothersome to scientists who have developed effective techniques in the laboratory. Such researchers have found it difficult if not impossible to obtain soil laced with PCB or to receive permission to carry out tests at a contaminated site.



These steel caissons create a laboratory on the floor of the Hudson River.

Daniel Abramowicz, manager of the bioremediation laboratory at General Electric (GE)'s Research and Development Center in Schenectady, New York, says that it has been "a research nightmare" to do experiments with PCBs, widely used in electrical power equipment until the 1970s. GE researchers have had to wait as much as a year or more to receive crucial permits and have been barred from exchanging samples. Although a procedure for PCB bioremediation has already performed successfully in several field trials, its progress is stymied by EPA regulations.

Scientists working with other toxic chemicals face similar problems only if they want to test their disposal techniques on large amounts of hazardous waste. Most toxic chemicals are exempt from onerous regulations if only small amounts of the substance are used, and the EPA is thinking

about raising the level at which researchers lose their exemption.

But PCB regulations, which are based on the Toxic Substances Control Act (TSCA) passed in the late 1970s, provide no such exemption. As a result, some researchers say that PCBs are far more tightly regulated than is necessary to protect the environment and human health.

For example, EPA rules dictate that scientists can work only with PCBs "originally packaged in...hermetically sealed containers of no more than five milliliters." The rule forces researchers to buy PCBs from commercial suppliers, although many of the compounds can be easily synthesized. The 'original packaging' standard also prevents researchers from working with waste-site samples. As a result, scientists often use their expensive, purchased PCBs to contaminate soil and water samples in the laboratory despite an abundance of the real thing.

Researchers can obtain exemptions from the packaging standard, but obtaining a permit is an expensive and arduous process that can take up to a year. "The permit process is designed more for industry," says Pamela Morris, who works on PCB bioremediation at the EPA laboratory in Gulfbreeze, Florida. "They wanted me to have an engineering-scale diagram of my technology unit. Basically, I'm doing microbiology in a flask." Morris was unable to do important experiments during a six-month wait for her permit.

Those with a permit face a morass of regulations on the transportation, handling and disposal of PCB samples. Even relatively small samples of PCB soil must be shipped in a hazardous waste tanker truck at a cost that may reach thousands of dollars. EPA rules also require detailed accounts of how the materials are used and disposed of. EPA officials acknowledge that exemptions for researchers working with small quantities of PCBs make sense, and they plan to release new regulations for comment by the end of the year.

The changes will come none too soon: last year, the agency found more than 34 million cubic yards of PCB-tainted material at 1,200 of the country's most polluted dump sites. But the only options now available are to burn the material or dump it in an approved landfill. Relaxing its mandate to prevent PCB pollution would allow EPA to help those trying to find more and better ways to repair the environmental damage.

Traci Watson

NEWS IN BRIEF

London. Anyone still interested in the research on cold fusion being carried out by chemists Martin Fleischmann and Stanley Pons are likely to find it harder than ever to get their hands on the data. Speaking last week in Southampton at the annual meeting of the British Association for the Advancement of Science, Fleischmann said that the pair now work for an "industrial concern", a Japanese think-tank called Technova Inc., and that they must take commercial considerations into account before discussing their work. However, Fleischmann would not say what might be commercially sensitive about the work, how it extends the experiments that caused a media storm in 1989 or how much money Technova was investing. He did say that a new laboratory was being built somewhere in France but refused to identify the site except to point out that it was "very nice" there. **I.M.**

Washington. A two-year congressional investigation of accounting practices at US research universities has revealed an arcane and capricious accounting system that permits large differences in what universities charge the government for the cost of supporting research.

Last week's report, by the General Accounting Office (GAO), does not contain any scandals similar to those uncovered in 1989 and 1990. Instead, it focuses on the two agencies — the Office of Naval Research (ONR) and the Department of Health and Human Services (HHS) — that share auditing duties for about 600 research universities. Those assigned to ONR can charge an indirect cost rate that averages one-fifth higher than those assigned to HHS. Overall, GAO calculates that US universities have overcharged the government by at least \$400 million. (The universities disagree, and have returned or withdrawn about \$16 million.)

GAO recommends that one agency oversee the system to make it more consistent. It also recommends that the government include university representatives in negotiations now under way. Although the report calculates the probable impact of various alternatives (one possibility — a flat 50 per cent indirect cost rate — would cut research indirect cost funding by \$84 million for a sample of 137 universities), it does not recommend a specific option. **C.A.**

Washington. A small bolt was apparently to blame for the failure of last month's joint US and Italian space mission (see *Nature* **358**, 529; 1992) to release a tethered satellite. Investigators at the US National Aeronautics and Space Administration (NASA) reported last week that the bolt, which was added to the tether structure earlier this year to strengthen it for launch, apparently interfered with the tether's winding mechanism and prevented it from extending the full 12 miles. The satellite ventured only 840 feet from the space shuttle. **C.A.**

Hurricane pounds Florida research facilities

More than 4,000 research primates were among the homeless after Hurricane Andrew destroyed parts of southern Florida last week. Winds as high as 160 miles an hour blew apart hundreds of cages holding more than 400 rhesus monkeys and baboons at one National Institutes of Health (NIH) facility and flattened a commercial primate breeding facility that contained an estimated 4,000 animals.

At the NIH facility, operated by the University of Miami, about 10 per cent of the animals were killed when the cages "exploded" and the high winds "basically turned the monkeys into birds", says Robert Rubin, vice president for research at the University of Miami. Those that survived soon faced other dangers. An erroneous radio account reported that the monkeys were infected with the HIV virus and were biting people, and Coast Guard troops at a nearby installation began shooting the monkeys out of the trees in which they had taken shelter. In the 30 minutes that it took officials at the NIH Perrine facility to correct the report (the animals were in fact bred to be virus-free) six monkeys had been killed.

A nearby commercial facility, owned by the Mannheimer Foundation, reportedly suffered even more serious damage. Telephone contact with the facility was impossi-

ble late last week, but Rubin says that initial reports indicated that the facility was heavily damaged and about 4,000 primates were on the loose, with an unknown number killed or seriously injured.

Collecting the animals turned out to be another challenge. National Guard troops were called in after reports of baboons foraging in the streets — citizens were warned not to approach them and to drop any food they might be holding to avoid fights with the animals—but chasing the animals proved futile. Eventually, facility staff hit upon the best solution: put some monkey chow in an open cage and wait. Capturing most of the animals turned out to be as simple as shutting the cage doors, Rubin says.

Other research facilities in Florida were less hard-hit by the storm. Florida International University suffered about \$5 million in damage to several buildings and the Coral Gables campus of the University of Miami also reported significant damage. The storm cost the University of Miami between \$1 million and \$2 million.

Perhaps the most significant impact of the storm is the time that researchers have lost. The University of Miami reported that at least 200 research staff had lost their homes. Although no one was killed, "we're all still shell-shocked", said one university

spokeswoman. Experiments at the Perrine facility are expected to be delayed for six months as staff attempt to rebuild the breeding stock.

Damage to marine research fleets based in southern Florida was largely averted by a decision to sail the vessels northwards. One 36-foot sailboat belonging to the University of Miami was carried some 1.5 miles inland by storm-driven waves. Several shore buildings associated with the marine and oceanographic programmes lost parts of their roofs and roof-mounted equipment, including antennae, satellite dishes and laser gear.

The National Aeronautics and Space Administration (NASA) has delayed the launch of its Mars Observer mission after an operation to protect the spacecraft from the hurricane accidentally contaminated it. NASA said that the launch, scheduled for 16 September, would be delayed by at least 10 days while technicians cleaned the craft. Tests revealed that technicians had inadvertently sprayed paint and metal fragments on to the craft when they piped impure nitrogen into the Titan 3 nose cone to guard against fires caused by storm-related electrical sparking. NASA officials said that the probe must be launched by 13 October to reach Mars on schedule in September 1993.

Christopher Anderson

Academy says US national parks need more research

Washington. Insufficient research at US national parks has led to dozens of mistaken decisions and ecological disasters, a US National Academy of Sciences (NAS) panel has concluded in a report* released last month. Decrying decades of poorly funded research, the NAS panel called for a separate category for science in the National Park Service budget as a way to end the practice of diverting research funds for other purposes.

The new review is at least the twelfth major report since 1963 to comment on the sorry state of research in the US national parks and the need for better science. But panel members hope that it will be more influential than its predecessors, in part because wildlife management decisions are increasingly the subject of litigation. Park Service officials, realizing that better research data can improve their chances of victory in court, have endorsed the conclusions of the NAS report.

The NAS report, like many of the previous reviews, points out that the Park Service has one of the smallest percentages of researchers of any federal agency. Only in the past year has the Park Service significantly increased its research staff, from 48 to 112 scientists, and even the current number represents only 3 per cent of its staff. The figure

is 8–10 per cent at other US land management agencies.

Panel members blamed many of the Park Service's most notorious instances of mismanagement of plants, animals and natural resources on its historic disregard of science. The report cites cases where well-meaning park rangers killed wolves, bears, coyotes and other natural predators while they fed more 'desirable' species such as deer and elk. Only when the growing number of protected animals began to eat their way through the parks did the managers realize the consequences of such interference with the natural equilibrium. Similarly, in the absence of research on the subject, park rangers for decades suppressed natural fires until it became obvious that small fires are essential in controlling undergrowth and preventing catastrophic forest fires.

Today, the Park Service must decide how to maintain natural equilibria while allowing increasing numbers of visitors to visit the parks. Some parks have already begun to limit attendance, but there is no research to suggest that response is the best approach.

The NAS report calls for an "explicit mandate" — including a distinct line-item funding category — for science in the Park Service. It also recommends a new position

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Are power plants the source of this haze over the Grand Canyon?

of chief scientist and an external scientific advisory board. **Christopher Anderson**

* *Science and the National Parks*, National Academy Press, 1992 (2101 Constitution Avenue NW, Washington, DC 20418).

Hawking on film: first-rate account of a first-rate mind

Los Angeles. This month, moviegoers across the United States will notice an unfamiliar face among posters featuring the likes of Madonna, Meryl Streep, Clint Eastwood and Marlon Brando. It is British cosmologist Stephen Hawking, starring from his wheelchair in an adaptation of his best-selling book, *A Brief History of Time*. Director Errol Morris has overcome the obstacles posed by his subject and the topic to present a compelling story of both the man and his science.

Stricken for the past 25 years with Lou Gehrig's disease, Hawking has lost the ability to speak or move and requires round-the-clock nursing care. Yet his disability has not stopped him from holding the Lucasian chair at Cambridge University, and the movie suggests a parallel between his life and his work. As with a collapsing black hole, Hawking has been imploding upon a personal core that centres on his mind. Yet just as a black hole signals its existence

by emitting radiation, Hawking continues to produce papers and work of high quality.

Hawking is the central character of the film, although his mother and others tell parts of his story. We see him grappling with questions of black-hole physics and cosmology, subjects that Morris presents with astonishing integrity and respect for the intelligence of his audience. Virtual particles and event horizons are part of the supporting cast, as are John Wheeler, Kip Thorne, Roger Penrose, Dennis Sciama and Don Page. The narration demands one's full attention; it contains such lines as "A black hole has an entropy; therefore, it has a temperature".

Hawking's disabilities are shown without mawkishness; rather than a poster child seeking pity, he is depicted as a strong-willed man quoting Samuel Johnson's famous line that knowing one is about to be hanged "concentrates the mind wonderfully". At the same time, Morris avoids a portrayal that, as some have already done, invites comparisons to Einstein. Hawking is a scientist of the first rank, Morris makes clear, but not a demigod.

Morris, who received critical and popular acclaim for *The Thin Blue Line*, a documentary that led to the release of a man

falsely convicted of murder, began work on this movie in 1989 at the request of Gordon Freedman of the Los Angeles-based Anglia Television. Morris, who has studied the history of science and attended lectures given by Wheeler at Princeton, found the offer attractive. Indeed, Harvard physicist Sidney Coleman, who served as science adviser for the film, says that he was impressed by Morris' grasp of basic quantum mechanics.

The film was shot in London in 1990 and edited in 1991 before making its US cinematic debut last month in Los Angeles. It cost \$3 million, a fraction of the usual Hollywood budget. It was shown last spring on British television.

Although the film succeeds in bringing complex scientific concepts to a general audience, it is not without its flaws. For instance, it would have been useful to hear Hawking state clearly that the fundamental problem in cosmology is to reconcile the prediction of general relativity that the universe began from a pointlike mathematical singularity with the impossibility (according to quantum mechanics) of such a singularity. Morris has tried, without much success, to explain some of the ideas Hawking has used to avoid the singularity, but he might have done better to present competing theories from such people as Alan Guth of the Massachusetts Institute of Technology.

However, the very fact that one can raise such points shows how different this film is from the usual fare. One should not criticise a Western shoot-'em-up for not meeting the standards set by Civil War historian Bruce Catton. Likewise, *A Brief History of Time* upholds the standards set by Hawking himself, which should be high enough to satisfy even the most discriminating viewer.

T. A. Heppenheimer

Dinosaur books get seal of approval from new society

The best dinosaur books will receive a seal of scientific accuracy from The Dinosaur Society, a Massachusetts-based group of scientists and enthusiasts formed to improve the public's understanding of dinosaurs. The society has identified more than 80 books deemed accurate by a scientific review panel,



and a handful of titles that it recommends with reservation because of relatively minor errors.

The society declined to release a list of the dinosaur books that flunked the

test, although it did single out *The Book of Dinosaurs and Other Prehistoric Animals* for giving inaccurate time and place information (locating *Thecodontosaurus* in Australia, a continent where it has never been found, for example), and misstating dinosaur sizes and weights. On the other hand, it gave top honours for accuracy and educational value to *The Illustrated Encyclopedia of Dinosaurs*, a book by David Norman, a Cambridge University palaeontologist.

The society was founded last year by dinosaur researchers and enthusiasts to increase the \$1 million now spent worldwide on dinosaur research each year. The society would like some of the money to come from companies selling dinosaur-related products to the public. The group also hopes to raise \$1 million a year for research through individual and corporate membership, donations, project sponsorships and product endorsements.

Christopher Anderson

The Dinosaur Society can be reached at PO Box 2098, New Bedford, MA 02741, or by telephone, 1-508-999-1168, and fax, 1-508-997-2469.

Chefs forswear genetically altered food

Restaurant-goers in the United States will soon see this symbol on the menus and windows of some of their favourite establishments as part of a campaign against genetically engineered foods. Ted Howard, director of the



Washington-based Pure Foods Campaign (spun off from Jeremy Rifkin's Foundation on Economic Trends), says the response to the campaign has been "overwhelming". Since June, some 1,500 chefs across the country have agreed not to serve genetically engineered food.

Although no such food is commercially available yet, Calgene Inc. of California received government approval in May to market a tomato made with a foreign gene that retards ripening. The fruit is expected to go on sale in about six months.

C.A.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Stephen Hawking

The debate on whaling

SIR — D. S. Butterworth's Commentary "Science and sentimentality" (*Nature* 357, 532; 1992) appeared during a meeting of the International Whaling Commission (IWC), events at which refute Butterworth's thesis. That thesis seems in essence to be that other scientists have "obfuscated", "moved goal-posts" and generally behaved unethically in acting as mere dupes of governments with secret "animal rights" agendas. He cannot substantiate it.

New rules for setting catch limits were adopted by a resolution opposed only by Norway. They take into account the real uncertainties, not imaginary ones as Butterworth would have *Nature* readers believe. Norway and Iceland have concluded that they will lead to unacceptably low catch limits and have taken unilateral action accordingly.

Most IWC members — nearly all of them ex-whaling countries — have been working for five years towards a management regime for renewed commercial whaling that fulfils the objectives of the convention under with the commission operates. "The real debate" is not "between some countries wishing to preserve industries . . . and others wanting [whales] classed as sacrosanct" but rather between that majority which insists that resumed commercial whaling must be such that there is little chance of further depletions and that already depleted stocks are restored to high, productive levels, and a minority of three (now two) whose primary interest is in taking profitable catches in the immediate future.

The majority has placed conditions on the implementation of the new rules: that more attention be given to devising ways of hunting and killing whales humanely, and that effective international monitoring of the application of the rules (including standards for scientific data) be instituted. Neither condition is new, but neither is acceptable to the Norwegian authorities. A statement by Norway's foreign minister on 22 July, calling minke whales "rats of the sea", illustrates how emotions are manipulated so that rational discussion of science and management is prejudiced.

There is a strong moral component of the debate, but it does not involve "animal rights". Rather, it concerns sustainability of use, inter-generational equity and the rule of international law, under which whales have a unique status as "highly migratory marine animals", whose exploitation and conservation are to be governed through the IWC.

As to the matter of MSY rates (a measure of potential productivity), the consensus was that a range of from 1 to 4

per cent is justified by the sparse data. Butterworth has recently been arguing that much higher rates are more "reasonable"; their adoption would, of course, give correspondingly higher immediate catch limits, and greatly increase the chances of inadvertent depletions of stocks.

Butterworth's claim that the term 'protected status' has been "skilfully" misused by "conservationists" to generate alarm of 'danger of extinction' is false. That the 'protection' level chosen in 1975, and reaffirmed in 1991, concerns catch-maximization is well-known to all who have followed the proceedings of the IWC.

Butterworth thinks the proposal by France for a whale sanctuary suggests "desperation among animal rightists". The French government is hardly one usually associated with 'animal rights' policies although it is well known for recent successful actions for the conservation of Antarctica. Most members praised the proposal. Seventeen of them co-sponsored a resolution, which the commission adopted, to ensure that it would be definitely considered next year. There is no evidence that any of those governments wish to abandon or distort science.

Sidney Holt

*International Fund for Animal Welfare,
Podere Il Falco,
06062 Città della Pieve (PG), Italy*

Funding errors

SIR — J. Katz (*Nature* 358, 10; 1992) has, with astonishing brevity, epitomized the errors almost universally committed in the funding of research: emphasis on the number rather than the quality of publications, funding of promising research projects rather than of individuals of proven ability, the existence of large research groups, where freedom of creative thought is absent, and the support of consensus science, with the end-result that new and original ideas are discouraged by the present system. I hope Katz's letter is read, and understood, by some of the higher echelons in the field of research funding.

I agree with both the points raised and the arguments brought forward by Katz on these issues, with one notable exception. He believes that fraud in research is committed because the stakes are high. But a person who can perpetrate fraud in research is simply unsuitable as a scientist. In the words of Plato: "Science without justice, and also without every other kind of virtue, is closer to cunning rather than wisdom". If we begin to rationalize lying in science we will end up in an "Erehwon" type of scientific community, where it will be *de rigueur* to express sympathy to perpetrators of scientific fraud.

Emmanuel T. Rakitzis

*Department of Biological Chemistry,
University of Athens Medical School,
Athens 11527,
Greece*

CO₂ emissions

SIR — Being reminded by leading articles in *Nature* that the recent Earth Summit in Rio de Janeiro that was being sponsored by the United Nations (UN) and remembering that there have long been complaints that the West shoulders an excessive proportion of the costs of running that organization prompted an idea that could provide a painless solution to the lack of a serious proposal for carbon dioxide regulation. If the 1992 UN contributions are recalculated per ton of carbon dioxide emissions documented for 1982, the contributions of North America and Europe are then

seen to be the same as the rest of the world (see table). Were the UN Earth Conference to agree that contributions be recalculated according to national 1992 carbon dioxide emissions, there could be no dispute as little would change but the world would see that our leaders have agreed on the importance of carbon dioxide emissions being centre stage in their deliberations.

R. T. D. Oliver

*Department of Medical Oncology,
London Hospital Medical College,
Turner Street, London E1 2AD, UK*

1. Marland, G. *The Prospect of Solving the CO₂ Problem through Global Representation*. National Technical Information Service Document TR039, 4-8 (US Department of Commerce, Virginia, 1988).

UN SUBSCRIPTIONS ADJUSTED PER TON OF CO₂ EMISSION

	Population (millions, 1985)	CO ₂ emission (x 10 ⁶ ton 1982)	1992 UN contribution	
			Total (£million)	per ton CO ₂ emission
North America	237	1,243	271	£0.22
Europe	250	1,280	280	£0.22
Rest of World	4,355	2,226	487	£0.22
Total	4,842	4,749	1,038	£0.22

Not entirely a backwater

SIR — Barbara J. Culliton tells us that Texas, now honoured with two Nobel prizewinners, used to be a scientific backwater (*Nature* 357, 623; 1992), but that does not fit with history. The zoology department of the University of Texas at Austin, for example, has been for decades a stronghold of *Drosophila* and population genetics.

Hermann J. Muller made his outstanding experimental contributions to the study of heredity in plants and animals there, culminating in 1927 with reproducing mutations and producing novel ones by irradiation, an achievement for which he was awarded the Nobel prize in 1946, nearly 40 years ahead of Michael Brown and Joseph Goldstein.

Before and after the Second World War, the same department was a favourite for postdocs and other working visits by population geneticists. It produced quite a few of its own. In 1966, simultaneously with the teams of Harry Harris in London and R. C. Lewontin in Chicago, W. S. Stone's group in Austin discovered, through protein gel electrophoresis, the high level of gene heterozygosity in natural populations, contrary to Muller's contention (*Proc. natn. Acad. Sci. U.S.A.* 56, 119–125; 1966).

Through the early 1970s, M. R. Wheeler at Austin put out the widely read *Studies in Genetics* series. Wheeler maintained the largest, healthiest and most varied existing collection of live fruitflies, meeting requests for stocks worldwide, and Stone was a key participant in a project for investigating Hawaiian drosophilids which, in the mid-1960s, received half a million dollars, an enormous amount for a research grant in biology at the time. Among *Drosophila* workers, the Stone group was famous for unparalleled preparations and pictures of giant salivary gland chromosomes.

The idea in the United States that anywhere but the east or west coast is a backwater strikes an all too familiar chord here in France, but several scientists from places other than Paris have nevertheless won Nobel prizes.

Georges Pasteur

*Ecole Pratique des Haute Etudes,
Place Eugène Bataillon,
34095 Montpellier 05, France*

Italian science

SIR — As members of the Science Committee of the Italian Space Agency (ASI) we feel obliged to rectify some of the statements made by Alison Abbott (*Nature* 357, 351; 1992).

It is obvious that the committee as a whole would welcome an increase of funds for basic science. The position expressed by the research minister in this respect is particularly important. It should be stressed, however, that "basic science" in ASI includes disciplines that are considered "applications" in the European Space Agency. ASI, at various levels, will have to engage in a careful discussion on how to apply the directives of the ministry.

We are concerned about the present functioning of the agency but we strongly feel that tilting at windmills will lead nowhere. So we have not changed our minds (as claimed in the article) and most of the members remain extremely critical of Professor Remo Ruffini's attitude, his personalistic approach and his poor management of the committees even in trivial matters.

It is probably unprecedented in science for an elected officer not to resign after a majority vote of no-confidence. We agree instead with Professor Giovanni Bignami that the internal mechanisms of the committee and of ASI itself (including the refereeing process) will require a careful, responsible re-examination.

Franco Pacini (*Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, 50125 Florence, Italy*); **Giuliano Boella** (*University of Milan*); **Ezio Bussolletti** (*Naval University of Naples*); **Marino Dobrowolny** (*Institute of Interplanetary Plasma, Frascati*); **Sergio Vetrella** (*University of Naples*)

The biter bit

SIR — A. P. R. Cooper¹ accuses Daedalus² of not having done his homework. It seems that Cooper has not done his either. The scheme proposed by Daedalus is not the well-known 'Lempel-Ziv-Welch' (LZW) algorithm³ for data compression. The LZW compression algorithm builds strings ('words') from input data, generating a good dictionary from most data with no previous knowledge of their contents. It is a variation of an algorithm by Lempel and Ziv⁴. In fact, a person knowing the current data-compression techniques would be tempted to divide Daedalus's scheme into two parts:

(1) The idea of representing each word (or phrase) in a text by a look-up, or pointer, into a static dictionary and (2) the idea of representing each pointer with a number of bits according to the probability of occurrence.

Part (1) has not got a name, as it is regarded as 'immediately obvious' and no one has (yet) claimed credit, but it could, with a bit of tolerance, be called a variation of the 'Lempel-Ziv' data compression algorithm. Part (2) is also a

well-known technique, but is generally used only on pure character data. The idea of using fewer bits for more common characters was proposed in the mid-1960s, and is currently taught at university level as 'Huffman-encoding'. This is not done in quite the way Daedalus proposed, as his algorithm cannot identify the length of the code for each word. To avoid this problem, Huffman-encoding is constructed so that no code introduces another code. An example can be given using Daedalus's scheme: assume that *the*, *a* and *an* are the most common words in *Nature*, in that order. Then *the* would be represented by binary 0, *a* would be represented by binary 1, and *an* would be represented by binary 10. If the program encounters the bits 10, it cannot tell whether this is supposed to mean 'a the' or 'an'.

Elvind Eklund

*Granveien 30,
N-1430 Ås, Norway*

1. *Nature* 358, 365 (1992).
2. *Nature* 358, 22 (1992).
3. Welch, T. A. *Computer* (June, 1984).
4. Ziv, J. & Lempel, A. *I.E.E.E. Transactions on Information Theory* (May 1977).

The reviewer reviewed

SIR — Book reviews, as is well-known, say as much about reviewer as reviewed; in some cases they may even serve as a form of therapy. But there is a limit, and Stuart Sutherland (*Nature* 357, 550; 1992) with his dismissive aside on the "crackpots" J. B. S. Haldane and J. D. Bernal, is, I suggest, the wrong side of it.

Bernal and Haldane, respectively primarily crystallographer and geneticist, but both indisputably polymaths, were not merely innovative experimental and theoretical researchers within the laboratory; both were inspired visionaries and writers about the goals and purposes of science in its widest social context outside the laboratory. No one who attended the recent centenary meeting for Haldane at the Science Museum, or who rereads today Bernal's *Science in History* or *The Social Functions of Science*, is likely to be in any doubt about this.

Although Bernal and Haldane are likely to be remembered long after Sutherland's sniping — and indeed the book he was ostensibly reviewing — are forgotten, it is a pity that *Nature's* pages should be thus occupied.

Steven Rose

*Department of Biology,
Open University,
Walton Hall,
Milton Keynes MK7 6AA, UK*

Last chance for British Ass?

The British Association for the Advancement of Science claims to have found a viable strategy for its own survival and for doing good works. It should be given the benefit of one last doubt.

THOSE reading reports (see, for example, page 5) from the annual meeting of the British Association for the Advancement of Science (BAAS), this year at Southampton, will be tempted to say, "But we've heard it all before." And, of course, they will be correct. This is certainly not the first or even the second occasion since the Second World War when the organization has shaken itself by the scruff of its neck and declared to itself, and to anybody else prepared to listen, that things would be different from then on.

Indeed, in the early 1970s, a grand reappraisal reached much the same conclusions as were advertised last week: there would be a periodical publication of some kind, more attention would be paid to matters of public importance (too often supposed to be, on that account alone, "controversial") and the organization would be more active in the year-long intervals between its annual jamborees. Since then, there have been fitful endeavours of that kind, but not such as to stir the world.

Will it be different this time? And does it matter whether the BAAS succeeds or fails in making a mark in the present world?

The short answers are "Don't know", and "Yes". But the second must be qualified by the recognition that not all nineteenth-century organizations can expect a licence to survive in the twenty-first. The long answer has to do with history.

The BAAS was founded in 1832, the year of the Great Reform Bill in which Britain took the first steps towards universal electoral suffrage. (A property qualification remained, and women had to wait almost a century for the vote.) The BAAS owes its existence to three influences: the ferment of interest in Britain in science and its application (Dalton was still active and the Industrial Revolution well under way), the interest of what are now called the chattering classes in the nature of the world (whence, eventually, this journal's existence) and the widespread (and correct) opinion that the Royal Society of London had become moribund.

For a good half-century, the BAAS was socially invaluable. So much is clear from the records of its early meetings, which were faithfully and often passionately recorded by the officials called "recorders" and then published in sufficient detail to be comprehensible. Its annual meetings were occasions when those of a philosophical temperament could learn what had happened to Dalton's atomic theory and what Lyell had been up to, when steel-masters could learn from metallurgists why it is that cast-iron has too much carbon to be malleable and when all and

sundry could learn what was happening elsewhere; overseas guests were enthusiastically received.

The highlights of the BAAS's history in the nineteenth century are probably overdone. The great debate on Darwinism between T.H. Huxley and Bishop Wilberforce is commonly cited as an illustration of the BAAS's knack of making public the great controversies in science. It may have mattered much more to those who attended the annual meetings that they were among the few occasions when the then handful of dedicated researchers rubbed shoulders for a week with those whose livelihoods or peace of mind hung on what they had to say. For the BAAS's saving grace, then as now, is that its members are not required to be qualified at anything in particular.

Sadly, in the twentieth century shoulder-rubbing acquired snobbish overtones. People took to wearing dinner-jackets (tuxedos) to the annual dinners of their specialist sections — and kept up the practice long after it should have been clear to them that productive researchers did not own such garments. People would stay at hotels that matched their station in life — and those who could not afford hotels would stay with friends, or come only for a day.

Yet at the beginning of this century, the annual meetings were often memorable for what researchers said in public for the first time. (There is good reason to believe that the Lorentz contraction of the length of a relativistically moving object is known as the "Lorentz-Fitzgerald contraction" only because Sir Oliver Lodge told an annual meeting that the Irishman Lorentz "has been teaching that to his students for years".) By 1945, with the emergence of a research profession in the universities and elsewhere, and the recognition that rapid publication is an obligation, they had become occasions active researchers were too busy to attend (except perhaps to give a talk and go away without reimbursement).

That does not mean that the annual meetings did not give pleasure to those who chose to attend them. Why would the audience have been there if it were otherwise? But, by the 1960s, the association behaved as if its function were to provide an annual occasion when upwards of 2,000 people (and 5,000 in a bumper year) could listen to the great, the good and the promising, taking in a few entertaining field excursions in the afternoons and an occasional dress-up dinner in the evenings.

The great crisis of the early 1970s was

occasioned by a shortage of funds and of goodwill. Then (as still) the annual meetings are made possible by the hospitality of universities, but universities were already short of funds. Worse, although there has always been a nucleus of administrative staff, the donkey-work of arranging scientific programmes was the responsibility of the recorders, who had ceased to be reporters and had become unpaid wheelers and dealers whose only reward was a modicum of the resemblance of power.

Twenty years ago, the response to crisis was typically English: committees were established, but were inadequately serviced. A handful of study groups was established to tackle issues of public importance. (The most successful of these, in which *Nature* had a hand, was an early study of the consequences of the then-new biology which had the now-Sir Walter Bodmer as its chairman and included as members Mrs Shirley Williams, the politician and Dr (now Lord) Owen, afterwards Foreign Secretary and now the European Bosnian peace negotiator.) But these initiatives ran into the sand for lack of funds and will.

Could it be the same again? Let us hope not. Although, or perhaps because, the BAAS is marked by its origins in the nineteenth century, it would be a great misfortune if it went to the wall. But there is a sense in which it would be preferable to fail in a brave venture at survival than to be paralysed by an excess of caution, as it has been in the recent past. That is why nobody should complain that the annual meeting has been rechristened a "festival" or why this year's occasion was enlivened by a demonstration by sceptics that it is, indeed, possible to walk on hot coals. If that brings a larger audience willing to listen to more serious business, so much the better.

The greater difficulty will be that of sustaining some kind of activity between festivals. Not, of course, that there is a shortage of good works to accomplish. In luckier times, the BAAS might well have become the defender of British science against repressive governments. (Its apology has usually been that its council is too diverse to have a single opinion.) Even now, there is the whole issue of public education in science crying out for attention. And what of the wider European dimension of British science? With a world as full of important issues as the present, the BAAS should find a surprisingly substantial part of the research community willing to give it the benefit of one last doubt.

John Maddox

Damage-limitation exercises

Deborah E. Barnes

SEVERAL rare genetic diseases, characterized by cellular hypersensitivity to DNA damage, confer an increased risk of cancer on people afflicted by them. In April, the cloning of complementary DNAs that correct the cellular defect in one of these diseases, Fanconi's anaemia, was announced¹. Now, on page 70 of this issue² we have a report of the same accomplishment for another of them, xeroderma pigmentosum. Both papers add to the roster of hitherto unknown genes involved in the cell's response to DNA damage.

Fanconi's anaemia (FA, with four genetic complementation groups) and xeroderma pigmentosum (XP, eight complementation groups), together with ataxia telangiectasia (AT) and Bloom's syndrome (BS), are a heterogeneous collection of recessively inherited human diseases; they are known as DNA-repair disorders because of the characteristic hypersensitivity to one or more DNA-damaging agents or mutagens. Cells from patients with the first three diseases are uniquely sensitive to the spectrum of DNA damage induced by DNA cross-linking agents, ultraviolet light and ionizing radiation, respectively; cells from patients with Bloom's syndrome appear to show a more generalized repair defect, being moderately hypersensitive to several agents that induce chemically distinct types of DNA damage. As aberrant repair of damaged DNA is reflected in an increased occurrence of chromosome abnormalities, these are also commonly referred to as chromosome-breakage syndromes.

So what do we know of the precise molecular defects that cause these syndromes? An intensive effort has been underway to identify the genes responsible by complementation of mutagen-hypersensitive cell lines. This strategy proved successful using rodent cells, which are amenable to the uptake of human genomic DNA, and genes that complement cells from XP group B (ref. 3) and XP group D (refs 4 and 5) were initially identified through complementation of repair-deficient rodent cell lines. But applying this technique of DNA-mediated gene transfer to human cells proved surprisingly difficult, and only the XPAC gene (XP group A complementing)⁶ and a candidate gene for one of at least two complementation groups of AT (ref. 7) have been identified directly by complementation of cell lines derived from patients with the respective disorders.

One obstacle to the gene-transfer approach in human cells has been their

recalcitrance to the stable uptake of DNA of high molecular weight. New complementary DNA expression cloning technology circumvents the rather hit-or-miss affair of stable and functional integration into the recipient genome by using Epstein-Barr virus-based vectors that replicate extrachromosomally in human cells⁸. Episomal maintenance of cDNA clones gives a high transfection efficiency and also facilitates their 'shuttling' back into *Escherichia coli* for further manipulation. These first-generation expression shuttle-vectors promise to be of widespread application for cloning cDNAs defective in any human cell line where a suitable selection strategy can be devised. Amongst the DNA-repair disorders, only the modest mutagen-hypersensitivity of BS precludes this approach, and definition of the molecular defect may instead come from pursuing candidate genes⁹.

The reports by Strathdee *et al.*¹, and Legerski and Peterson², herald the application of the cDNA expression strategy to two DNA-repair disorders. They identify cDNAs that specifically complement the hypersensitivity to DNA crosslinking agents of an FA-C cell line¹, and reduce the cellular ultraviolet-sensitivity of XP-C to normal levels². The sequence of the FACC cDNA gives no clues to the function of the encoded protein, but an inherited missense mutation in the only cell line representative of FA group C provides evidence for an altered FACC gene as the primary defect. A single base deletion in one allele of the gene in two previously unclassified FA cell lines also places these firmly in group C. How, though, can a single mutant allele result in the FA phenotype when inheritance of the disease is recessive? It must be assumed that the 'good' allele in each case is not expressed. This phenomenon is not without precedent, and has also been seen in an XP-B cell line where all detectable transcripts were derived from just one (mutant) allele of the XPBC gene³.

The molecular basis of the XP-C defect is not yet clear, but Legerski and Peterson² found no XPCC transcripts in four out of five XP-C cell lines examined, which supports the hypothesis that XPCC is the defective gene. Indeed, the amino-acid sequence of the XPCC cDNA shows significant similarity to the product of the *RAD4* DNA-repair gene of budding yeast. This fits into an emerging pattern, in which XP cDNAs share homology with yeast *RAD* proteins that are required for incision of the DNA

backbone during nucleotide excision repair. It is no surprise that essential enzymes of DNA metabolism have been conserved through evolution, and further insights into the workings of the cell's DNA-repair machinery will no doubt be gained through studies of the yeast *rad* mutants.

Meanwhile, the advent of a cell-free system that can mediate DNA nucleotide excision repair in human cell extracts allows direct functional analysis of the products of DNA-repair genes, as has been demonstrated with the XPAC protein¹⁰. In the clinical context, identification of the FACC and XPCC genes will make it easier to assign cases of FA and XP to genetic complementation groups and improve diagnosis.

Tumorigenesis is a multistage process which involves the inheritance and/or accumulation over time of mutations in proto-oncogenes, and mutations accompanied by mitotic recombination/chromosome loss in tumour-suppressor genes¹¹. The stability of the genome depends in turn on the concerted action of many gene products involved in DNA replication, DNA repair and genetic recombination. As the inherited DNA-repair disorders are associated with predisposition to cancer, genes such as FACC and XPCC have been termed tumour-preventing^{3,12}, their aberrant functioning setting the accumulation of somatic mutations on fast-forward, with malignancy being the almost inevitable outcome of an excessive mutational load.

However, the DNA-repair disorders do not fit comfortably into this simple scheme: not all repair-deficient disorders confer an increased risk of cancer¹³; there is a strong bias (except in BS) towards development of particular tumour types¹⁴; and the cancer risk may be a secondary consideration (only FA patients surviving early bone marrow failure are then predisposed to myeloid leukaemia). Furthermore, a predisposition to cancer, the DNA-repair disorders display a bewildering array

1. Strathdee, C. A., Gavish, H., Shannon, W. R. & Buchwald, M. *Nature* **356**, 763-767 (1992).
2. Legerski, R. & Peterson, C. *Nature* **359**, 70-73 (1992).
3. Weeda, G. *et al.* *Cell* **62**, 777-791 (1990).
4. Weber, C. A., Salazar, E. P., Stewart, S. A. & Thompson, L. H. *EMBO J.* **9**, 1437-1447 (1990).
5. Flejter, W. L., McDaniel, L. D., Johns, D., Friedberg, E. C. & Schuit, R. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 261-265 (1992).
6. Tanaka, K. *et al.* *Nature* **348**, 73-76 (1990).
7. Kapp, L. N. *et al.* *Am. J. hum. Genet.* **51**, 45-54 (1992).
8. Yates, J. L., Warren, N. & Sugden, B. *Nature* **313**, 812-815 (1985).
9. Barnes, D. E., Tomkinson, A. E., Lehmann, A. R., Webster, A. D. J. B. & Lindahl, T. *Cell* **69**, 495-503 (1992).
10. Robins, P., Jones, C. J., Biggerstaff, M., Lindahl, T. & Wood, R. D. *EMBO J.* **10**, 3913-3921 (1991).
11. Fearon, E. R. & Vogelstein, B. *Cell* **61**, 759-767 (1990).
12. Wood, R. D. & Lindahl, T. *Nature* **348**, 13-14 (1990).
13. Lehmann, A. R. *Cancer Res.* **7**, 82-103 (1987).
14. Cairns, J. *Nature* **289**, 353-357 (1981).

of clinical symptoms, including immunodeficiencies, neurological problems, skeletal abnormalities and altered growth, which are characteristic for a particular disorder.

Having identified the genetic defects underlying several of the DNA-repair disorders, we clearly need to look at the functioning of both normal and mutant gene products to get a clearer view of the clinical picture. Insights into the physiological repercussions of a DNA-repair defect will be gained by analysis

of animal models generated using homologous gene disruption and transgenesis in the mouse — just as isolation of human DNA-repair genes was first achieved by complementation of mutant rodent cell lines, knowledge of the relationship between defects in these genes and the disease phenotype may also come from rodent models. □

Deborah E. Barnes is in the Imperial Cancer Research Fund Clare Hall Laboratories, South Mimms EN6 3LD, UK.

PALAEOCLIMATOLOGY

The algal breath of life

Robert Riding

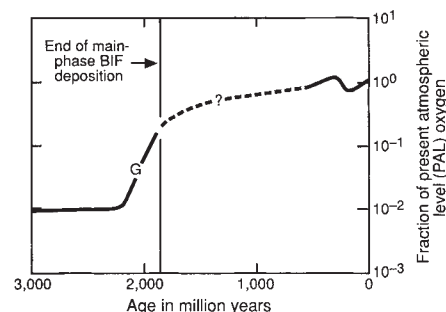
DISCOVERY of the earliest known alga, described in *Science* by Tsu-Ming Han and Bruce Runnegar¹, pushes the record of eukaryotes back a further 300 million years (Myr), from 1,800 Myr (ref. 2) to about 2,100 Myr. This was a critical period in the evolution of the Precambrian oceans when conditions appear to have been changing from anaerobic to aerobic. Algae require oxygen, so their existence at this juncture adds intriguing information to estimates of just when, and by how much, oxygen levels rose on the early Earth.

A central consideration in discussions of oxygen evolution during the Precambrian (that is, before 550 Myr) has been the significance of banded iron formations (BIFs). These large deposits of iron oxides, which today are major reserves of iron ore, are witnesses to a radical step in atmosphere-hydrosphere evolution. When they formed about 2,000 Myr ago they provided a sink for newly formed oxygen that was used in the oxidation of ferrous iron in the oceans. Only when the supply of ferrous iron ran out was it possible for concentrations of free oxygen to begin to build up in the oceans and atmosphere. This emphasis on the importance of BIFs is one aspect of a seminal biogeochemical synthesis of early life and environment formulated by the late Preston Cloud³. The model predicts that eukaryotes — which are obligate aerobes — would not have appeared until after the end of the main phase of BIF deposition. Up to that time the environment was anaerobic.

It comes as a surprise, therefore, that the fossil — *Grypania* — reported by Han and Runnegar does not postdate the main phase of BIF deposition (see figure). Not only is *Grypania* contemporaneous with BIFs, but it actually occurs within one of the principal BIF accumulations near Marquette, Michigan, close to the southern shore of Lake Superior. Han and Runnegar regard

Grypania as an alga and estimate that it required at least 1% of the present atmospheric level (PAL) of oxygen to survive. If this is correct then it conflicts with the view that BIFs reflect global anoxic conditions.

Grypania is a simple, but relatively



Oxygen rise during the past 3,000 Myr estimated from geochemical and fossil evidence (after Fig. 13, ref. 7; see also Fig. 2, ref. 9). The rapid rise around 2,000 Myr is inferred from fossil soil and other geochemical data, and is mainly attributed to cyanobacterial photosynthesis. Han and Runnegar's discovery of *Grypania* (G) at 2,100 Myr pushes back the age of the previous oldest eukaryotes and of the previously oldest known *Grypania*. Note that the shape of the oxygen curve between about 1,850 and 500 Myr is uncertain.

large, spirally coiled filament about 1 mm wide and up to 90 mm long. The specimens found by Han and Runnegar are abundant, but they are only carbonaceous films; no cells are preserved, and there are few details apart from the external size and shape of the filament. Despite the limitations imposed by this meagre evidence, Han and Runnegar not only interpret *Grypania* as a eukaryote, but also as an alga, and compare it with the extant dasycladalean green alga *Acetabularia*.

Although specimens this old have not been described before, *Grypania* is fairly well-known from younger Precambrian

rocks. Late last century, Charles D. Walcott considered *Grypania* from Montana to be a trace fossil (trails left on sediment by metazoans). Subsequently it was reinterpreted by Cloud as a "probable alga", a view supported by Malcolm Walter and co-workers⁴, although others had serious reservations. Only with the discovery of better-preserved material in India and China did it become clear that *Grypania* filaments were not only spiral in form but were also cylindrical in cross-section and possessed transverse markings⁵.

Han and Runnegar argue that the size, complexity and rigidity of *Grypania* preclude a bacterial affinity but are consistent with it being an alga. This is plausible. Nonetheless, direct links between *Grypania* and either green algae in general or dasycladaleans in particular seem tenuous. Perhaps one is prejudiced here by the expectation that red algae should appear before greens, a view supported by both phenotypic and molecular phylogenies. The earliest red alga currently known is a well-preserved bangiophyte no older than 1,300 Myr⁶.

Although it would be of interest to know the precise affinities of *Grypania*, the real significance of the discovery lies simply in the likelihood that it is a eukaryote. The emergence of eukaryotic organization paved the way for the divergence of algae, other plants, fungi and animals. It was one of the most significant thresholds in the development of life, and there seems little doubt that Han and Runnegar have pushed it back by 300 Myr. Even by the extravagant time standards of the Precambrian this is substantial, and is especially important because it places a eukaryote firmly within the timing of BIFs.

As a eukaryote, *Grypania* would have required oxygen for respiration. From its size and cylindrical shape, Runnegar⁷ estimated that it would have needed 1–10% PAL oxygen to survive. Such biologically based estimates provide useful minimum values to inform the vigorous debate concerning oxygen levels

- Han, T.-M. & Runnegar, B. *Science* **257**, 232–235 (1992).
- Zhang Zhongying *J. Micropalaeont.* **5**, 9–16 (1986).
- Cloud, P. *Paleobiology* **2**, 351–387 (1976).
- Walter, M. R., Oehler, J. H. & Oehler, D. Z. *J. Paleont.* **50**, 872–881 (1976).
- Walter, M. R., Du Rulín & Horodyski, R. J. *Am. J. Sci.* **290A**, 133–148 (1990).
- Butterfield, N. J., Knoll, A. H. & Swett, K. *Science* **250**, 104–107 (1990).
- Runnegar, B. *Palaeogeogr. Palaeoclimat. Palaeoecol.* **97**, 97–111 (1991).
- Towe, K. M. *Palaeogeogr. Palaeoclimat. Palaeoecol.* **97**, 113–123 (1991).
- Kasting, J. F. *Palaeogeogr. Palaeoclimat. Palaeoecol.* **97**, 125–131 (1991).
- Holland, H. D., Feakes, C. R. & Zbinden, E. A. *Am. J. Sci.* **289**, 362–389 (1989).
- Holland, H. D. & Beukes, N. J. *Am. J. Sci.* **290A**, 1–34 (1990).
- Beukes, N. J. & Klein, C. in *The Proterozoic Biosphere* (eds Schopf, J. W. & Klein, C.) 147–152 (Cambridge University Press, 1992).
- Anbar, A. D. & Holland, H. D. *Geochim. cosmochim. Acta* **56**, 2595–2603 (1992).

throughout the Precambrian, and particularly during the inferred rise around 2,000 Myr^{8,9}. Geochemical studies of iron retention in fossil soils suggest that atmospheric oxygen may have risen from 1% to 15% PAL between 2,200 and 1,900 Myr^{10,11}.

This combination of evidence from fossil algae and fossil soils makes it increasingly difficult to regard BIFs as indicators of overall conditions near the Earth's surface. The problem is intensified by revised age estimates suggesting that the main phase of BIF deposition ended around 1,850 Myr¹², substantially later than the 2,000 Myr that Cloud had in mind.

The coexistence of an obligate aerobe, such as *Grypania*, and BIFs supports the view that BIFs accumulated under a mildly oxygenated atmosphere in the early Proterozoic (around 2,100 Myr),

although they may well have formed under anoxic conditions before then¹³. This could spell the end of BIF-dominated models of oxygen build-up in the early atmosphere. Nonetheless, there remains broad consistency between the occurrence of *Grypania* and oxygen levels inferred from other geochemical evidence for the period around 2,100 Myr (see figure). The cat will really be put amongst the pigeons, however, if fossil discoveries extend the eukaryote record back much beyond about 2,200 Myr, into what is still widely perceived to have been an essentially anaerobic world. □

Robert Riding, Department of Geology, University of Wales College of Cardiff, Cardiff CF1 3YE, UK, is currently at the Institute for Geology and Palaeontology, University of Tübingen, Germany.

GLACIATION

Explosive start to last ice age

V. Ramaswamy

THE massive volcanic eruption at Toba, in the Indian Ocean, 73,500 years ago, may have accelerated the onset of the last ice age by loading the stratosphere with aerosols and shadowing the Earth, Rampino and Self suggest on page 50 of this issue¹.

The effect of volcanic eruptions on climate was recognized at least as far back as the time of Benjamin Franklin. The eruption at Tambora in 1815 was followed by the 'year without a summer'; and more recently we have been able to discern the effects of Agung (1963), El Chichón (1982) and Mount Pinatubo (last year). Toba was a far more massive volcano than any of these² and probably injected 5 times as much material into the stratosphere as Tambora, and perhaps 50 times more than El Chichón³. The oxygen isotope record and other palaeoclimatic evidence cited by Rampino and Self shows that the eruption occurred when the Northern Hemisphere climate had already begun a transition that eventually led to the last great ice age. The resulting 'volcanic winter' at a critical time during the interglacial-glacial transition may have sped up the climate change already underway.

A super-eruption such as Toba would have injected vast quantities of material well into the stratosphere to altitudes exceeding 25 km. Although dust from the eruption would have fallen out within a few months, the sulphur volatiles would have led to the formation of longer-lived sulphate aerosols¹. The evolution of the particulates and their

dispersal worldwide would depend on such factors as the amount of sulphur gases and water vapour emitted, microphysical processes of particle growth and coagulation⁴ and changes in the wind patterns. The authors estimate that the stratospheric aerosol optical depths could have been as high as 10 (extinguishing direct sunlight e¹⁰-fold) during the initial stages and would have remained above 0.1 for several years. In contrast, the peak optical depths from the recent El Chichón and Mount Pinatubo eruptions are estimated to be less than 0.5, and that for Tambora³ may have been 1. The substantial optical depths due to the Toba eruption would have caused a significant reduction in the sunlight levels for several years following the eruption, leading to a 'volcanic winter' — a cooling of about 3–5 °C in the Northern Hemisphere during the year following the eruption. For 2–3 years in the high latitudes the summers may have been as much as 10–15 °C cooler, leading to perennial snow cover over the central plateau of Quebec and Labrador.

The Toba aerosols would have faded over the course of several years and the blocking of sunlight would eventually cease. Would that have been the end of Toba's influence? With a significant positive feedback mechanism in operation, the climate effects may have persisted far longer. The authors point out just such a mechanism: the volcanic winter could have caused an increased growth of snow-cover and sea-ice which, by reflecting more sunlight, would have further cooled the planet. This positive

feedback could have accelerated the transition to glacial conditions.

Palaeoclimatic evidence indicates that the stage 5a–4 transition from interglacial to glacial conditions had already begun at the time of the eruption. The Earth's slowly varying orbital parameters then gave low levels of summer insolation in the high northern latitudes, CO₂ levels in the atmosphere were decreasing, sea levels were falling and cooling was in progress. Indeed, the geophysical changes accompanying the climatic transition, by changing the way the Earth's crust was loaded with water and ice, may well have triggered the volcanic eruption⁵. Rampino and Self point out that, coincident with the Toba eruption, there appears to have been a more rapid decrease in the North Atlantic surface temperatures and global sea levels, which they tie to the volcanic winter, particularly through the retention of the high-latitude ice sheets during summer.

It would be interesting to examine the palaeoclimatic evidence at the time of other super-eruptions and determine whether the volcano-climate interrelationship can be generalized. Another challenge lies in using general circulation models of the atmosphere to investigate the phenomenon of increased ice-sheet growth and perennial snow-cover under conditions of volcanic winter, especially in the high northern latitudes, and to see whether Rampino and Self's proposed mechanism could have accelerated the interglacial-glacial transition.

The consideration that the severity of a volcano's effect upon climate depends on the state of the climate system at the time of eruption prompts speculative questions — would glaciation not have occurred or would it have been less rapid had Toba not erupted? There is also the suggestion¹ that during the reverse, glacial-interglacial transitions, volcanoes may actually retard the warming of the climate. Rampino and Self even suggest that the linkage between the Earth's orbital parameters (the Milankovitch cycles), timing of volcanic eruptions and prolonged climate effects may be more than just a coincidence. Indeed, the link may be physically distinct, with climatic transitions modulating volcanic activity, and the subsequent volcanic-winter conditions modifying the climate itself. □

V. Ramaswamy is at the Geophysical Fluid Dynamics Laboratory, Princeton University, PO Box 308, Princeton, New Jersey 08542, USA.

1. Rampino, M. R. & Self, S. *Nature* **359**, 50–52 (1992).
2. Chesner, C. A. et al. *Geology* **19**, 200–203 (1991).
3. Rampino, M. R., Self, S. & Stothers, R. C. A. *Rev. Earth planet. Sci.* **16**, 73–99 (1988).
4. Pinto, J. P., Turco, R. P. & Toon, O. B. *J. geophys. Res.* **94**, 11165–11174 (1989).
5. Rampino, M. R., Self, S. & Fairbridge, R. W. *Science* **206**, 826–828 (1979).

Twilight of the pygmy hippos

Jared M. Diamond

WHO or what caused the late-Quaternary extinctions of many large mammals — were humans responsible, or climate change¹? For a long time, the best-known examples were provided by the Americas and Australia, to which more recently has been added every palaeontologically explored island of the Pacific. Surprisingly, European biologists and archaeologists have tended to ignore examples much closer to home, on the islands of the Mediterranean² and western Indian Ocean³. Discoveries about the extinct pygmy hippos of Cyprus⁴⁻⁷ and Madagascar⁸ are now yielding tantalizing new insights into the date, speed and cause of the extinctions.

Mediterranean islands such as Cyprus, Crete, Corsica and Sardinia are surrounded by such deep water that they remained unconnected to Europe even at Pleistocene times of low sea level. All were reached (by swimming or rafting) by flightless mammals, which then underwent striking evolutionary changes in body size⁹⁻¹². In general, big mammals became smaller, while little ones became larger. The islands developed pygmy elephants, hippos and deer, and even a paradoxically named pygmy giant deer (*Megaceros*). Leg bones indicate that these dwarfs had become slower-moving than their mainland ancestors, as would be expected given the absence of large predators⁹. Side by side with the dwarfs lived monsters: big shrews and voles, as well as giant dormice, owls, swans, lizards and tortoises.

All of these insular dwarfs are known only from subfossil bones, some of which were formerly believed to be remains of early Christian martyrs. The bones date to no more recently than the early or middle Holocene, which began about 10,000 years ago. Human settlement of Mediterranean islands did not begin until the early Holocene, and (as elsewhere in the world) this approximate coincidence between megafaunal extinction, human settlement and the end of the Pleistocene cries out for clarification. Were the extinctions due to climate and habitat changes at the Pleistocene/Holocene boundary, or to a *Blitzkrieg* of human hunting or habitat changes caused by humans and their domestic animals?

Until recently, the earliest known human occupation of Cyprus was by Neolithic herders and farmers¹¹. Their sites are marked by remains of stone houses, stone bowls and the bones of the sheep, goats and pigs that they brought with them. The sole remains of the extinct megafauna at these sites are two

debated scraps of hippo bones. Conversely, none of the previously known Cypriot sites with abundant bones of pygmy hippos and elephants yielded signs of human presence. That would seem to mean that the dwarfs succumbed to climate changes just before human arrival.

Discoveries⁴⁻⁷ at the Cypriot site of Akrotiri-Aetokremnos¹³ reverse this conclusion, however, and at last support the view that hippos and humans coexisted (even if only briefly). Simmons and Reese have unequivocally established the presence of humans at the site by identifying hearths, shell beads and stone artefacts very distinct from the Neolithic implements found on the island. There is no trace of domestic animals, suggesting that the people were hunter-gatherers.

The presence of the pig-sized hippos is also unequivocally established — their bones occur in prodigious quantities, and over 240,000 of them are known from at least 200 individuals. Along with the hippo bones are remains of lots of shellfish, crabs and sea urchins, many big birds, and at least three pygmy elephants, plus turtles and snakes. The bones are disarticulated and many are burnt, indicating that the people who left the stone tools cooked the animals that left the bones.

The age of the site is well supported by carbon-14 dates on various materials clustering around 8500 BC (uncalibrated). This makes the site 1,500–2,000 years older than Cyprus's Neolithic horizon, and possibly the oldest established site for human occupation of any remote Mediterranean island¹⁴. Interestingly, the proportion of hippo bones decreases, and the proportions of bird and mollusc bones increase, from deeper to shallower levels at the site, but both levels have indistinguishable carbon-14 dates. This implies that these first hunters to reach Cyprus quickly exterminated the pygmy hippos, which would have been conspicuous prey, easy to locate and kill down to the last individual.

It remains to be seen whether this human role in the extinction of Cyprus's pygmy hippos will apply to the miniaturized megafaunas of other Mediterranean islands. For Corsica and Sardinia, Vigne¹⁰ has summarized evidence that megafaunal species were alive when humans arrived, and vanished soon afterwards. My own guess is that all megafaunal species, on all remote Mediterranean islands, that survived into the Holocene, will prove to have become extinct through human agency.

Parallels and contrasts are presented by Madagascar's pygmy hippos, on which the Sun set long after their Cypriot brethren had vanished. Madagascar has yielded subfossil bones of giant lemurs and elephant birds, and of semi-pygmy (cow-sized) hippos³. The island has generally been believed to have been settled by humans around AD 500, and all these species were apparently extinct by the time that Europeans began describing Madagascar's animals in the seventeenth century. Very few human-modified bones of extinct megafauna are known from Madagascar, but MacPhee and Burney⁸ have now identified seven Madagascar hippo femurs bearing marks of having been cut and then snapped by humans. The cuts were apparently made by iron tools at a time when the bones were still fresh. Accelerator mass spectrometry carbon-14 dates on collagen from the bones fall between AD 0 and 300, pushing back the date of human occupation of Madagascar by a few centuries.

A possible contrast between Madagascar and Cyprus is the hint that the Cyprus megafauna succumbed quickly, whereas some of Madagascar's giant lemurs were still alive in the thirteenth century AD. The possibly slower pace of extinction on Madagascar may reflect its much greater area, and also a difference between the economies of the first Madagascar and Cypriot settlers¹⁵. The discoverers of Cyprus may have been full-time hunter-gatherers relying heavily on hippos for food, whereas Madagascar's first settlers were probably herders, farmers and fishermen for whom hunting was no more than a pastime. However that may be, the message of the new work is that zoology departments in Europe can dig into the reasons for extinctions on islands close to home. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.

1. Martin, P. S. & Klein, R. G. (eds) *Quaternary Extinctions* (University of Arizona Press, Tucson, 1984).
2. Martin, P. S. in *Quaternary Extinctions* (eds Martin, P. S. & Klein, R. G.) 354–403 (University of Arizona Press, Tucson, 1984).
3. Dewar, R. E. in *Quaternary Extinctions* (eds Martin, P. S. & Klein, R. G.) 574–593 (University of Arizona Press, Tucson, 1984).
4. Simmons, A. H. *Antiquity* **65**, 857–869 (1991).
5. Simmons, A. H. in *Perspectives on the Past* (ed. Clark, G.) 282–304 (University of Pennsylvania Press, Philadelphia, 1991).
6. Reese, D. S. *Cyprus View* **6**, 50–53 (1992).
7. Simmons, A. H. & Reese, D. S. *Archaeology* (in the press).
8. MacPhee, R. D. E. & Burney, D. A. *J. arch. Sci.* **18**, 695–706 (1991).
9. Reese, D. S. *Field Mus. nat. Hist. Bull.* **60**, 22–29 (1989).
10. Vigne, J.-D. *Atti Convegno Lincei* **85**, 369–392 (1990).
11. LeBrun, A. et al. *L'Anthropologie* **91**, 283–316 (1987).
12. Sondaar, P. Y. *Nat. Hist.* **95**, 50–57 (1986).
13. Simmons, A. H. *Nature* **333**, 554–557 (1988).
14. Cherry, J. F. *J. Mediterranean Arch.* **3**, 145–221 (1990).
15. Wright, H. T. et al. *Taloha* **11**, 121–145 (1992).

A matter of give and take

Richard W. Carlson

THE chemical composition of molten rock that erupts at the Earth's surface differs from that of the source rock, reflecting the influence of a series of processes that alter the liquid as it forms, migrates and emerges. The least understood of these processes occurs during transport of the magma from its source region in the Earth's interior to its point of eruption. If melt migration takes place through mechanisms that allow quick, large-volume flow — through fractures (dikes) or as rising bubble-shaped masses (diapirs) of magma, say — the melts may not interact significantly with surrounding rock. On the other hand, should melts migrate by slow transport along pore spaces between grains in a solid rock, they may be affected dramatically by the changing pressures, temperatures and compositions of the wall rocks they pass through. N. Takahashi (on page 52 of this issue), E. Takazawa *et al.* (page 55) and P. B. Kelemen *et al.* (*Nature* **358**, 635–641; 1992) each describe field evidence for the nature of melt segregation and migration and the role this step plays in modifying the composition of magmas and the rocks through which magmas pass to reach the surface.

Most source rocks for melts in the Earth's interior are composed of many different mineral phases. Consequently, melting of a rock within the Earth's mantle occurs over a range of several hundred degrees in temperature. From the point at which the first melt is produced to when the last mineral is consumed, different mineral phases enter into the melt in varying proportions, leading to substantial changes in melt composition with degree of partial melting. Because silicate melts generally are less dense than the solids they are generated from, buoyant separation of melt

from the host rock will take place before complete melting can occur. This leads to a spectrum of possible melt compositions that depends primarily upon the degree of melting reached before the solid and melt become separated.

Similarly, cooling of a magma in a sub-surface storage chamber leads to the crystallization of compositionally distinct mineral phases. Removal of the crystallizing minerals, by gravitational settling or by other means, can lead to extensive changes in magma composition.

The effects of partial melting and crystallization in magma chambers have been studied for decades, both in the laboratory and by examination of the eruptive products of volcanic systems. What has proven more difficult to investigate is the compositional changes imposed upon a magma during its passage through the Earth. Of critical importance in determining the significance of melt migration for chemical modification of a magma is the question of how long a magma spends in porous flow before it creates and fills large dikes that allow rapid transport with minimal contact with surrounding rock.

In this issue, Takahashi describes an example of this transition from micro- to macro-permeability from the Horoman complex of Japan, a slice of mantle emplaced into the crust by a plate collision in the past. As illustrated in the figure from Takahashi, flow of high-temperature melt through a fracture instigates melting of the surrounding rock. As the crack enlarges, suction pressure from the growing fracture draws melt from the surrounding rock. The result is a concentric ring around the fracture that shows increasing depletion in components removed by the melt with increasing proximity to the initial flow channel. In the examples studied by

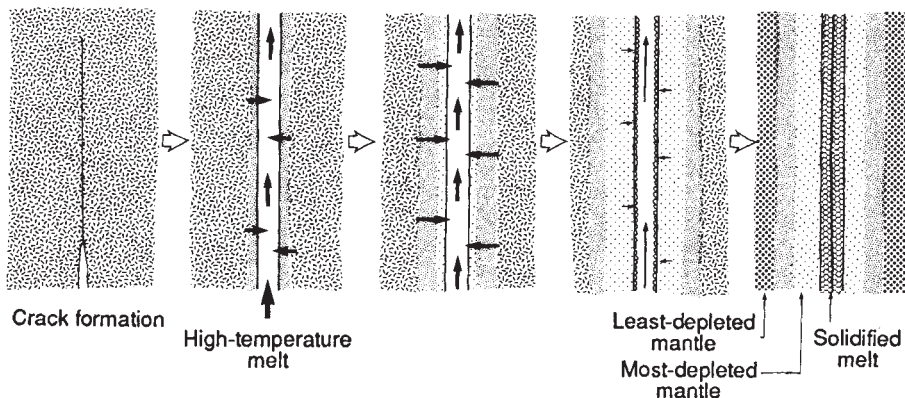
Takahashi, the "depletion zones" around the dikes extend from tens of metres to a few hundred.

Residues of partial melt removal are relatively easy to identify. In the mantle, the melting process selectively removes certain minerals, first Ca-rich pyroxene then Ca-poor pyroxene, ultimately leaving nothing but the Mg-rich mineral olivine. The chemical effect of this process is the removal from the residue of 'incompatible' elements, those elements that because of large ionic radius (for example, Rb, Cs, Ba, light-rare earth elements) or high ionic charge (Ti, Nb, Zr) do not fit comfortably into the crystal structures of mantle minerals, particularly olivine. Mantle rocks rich in olivine, deficient in pyroxenes, and depleted in incompatible elements are quite common, especially in the ocean basins. The mineralogical and chemical characteristics of these rocks fit well the simple model of them being the residues left behind by extraction of partial melt.

Another large group of mantle rocks, however, are rich in olivine, deficient in high-abundance incompatible elements such as Ca, yet have higher than expected abundances of incompatible trace elements. The trace-element study of Takazawa *et al.* identifies the Horoman complex as such a body. In spite of the simple trends in major-element abundances and mineralogy which led to Takahashi's model for melt extraction, these rocks have increasing relative abundances of highly incompatible trace elements with increasing major-element and mineralogical depletion.

Takazawa *et al.* invoke the long-held model that such contradictory trace-element and major-element characteristics reflect first a melt removal event like that outlined by Takahashi followed by infiltration of the residue by a melt that was highly enriched in incompatible trace elements. Interaction of this second melt with the depleted residues of the first event will then re-enrich the residue in incompatible trace elements, but, if the infiltrating melt is small enough in volume, it will not greatly alter the major-element and mineralogical characteristics of the rock.

This type of interaction is likely to happen if porous flow occurs over long distances. Porous flow will cause melts to be exposed to changing temperatures and pressures that will drive them out of chemical equilibrium with the surrounding rock, even if that rock does not vary in composition. As described two weeks ago by Kelemen *et al.*, this process causes the selective dissolution of Ca-rich pyroxene and precipitation of Ca-poor pyroxene and olivine in the solid residue of the reaction. The removal of Ca-rich pyroxene and the precipitation of olivine in the residue mimics the



Melt withdrawal caused by suction associated with melting along a growing fracture: the end result is a concentrically zoned body showing decreasing melt depletion going away from the original magma flow channel. (From N. Takahashi *Nature* **359**, 52–55; 1992.)

effects of simple melt withdrawal, but the deposition of Ca-poor pyroxene leads the residue to be overly rich in silicon compared with that expected for melt withdrawal.

Anomalous Si enrichment in otherwise depleted mantle rocks is a common feature of the deep mantle samples brought to the surface from beneath continents by certain types of explosive volcanic eruptions. If the explanation proposed by Kelemen *et al.* for the anomalous Si and incompatible-element abundances of many of these samples of deep subcontinental mantle is correct, the evidence

suggests that much of the deep, stable mantle beneath continents has been modified chemically by the passage of melts. Thus, interaction with passing melts may be a process of comparable importance to partial melting and crystallization in magma chambers in determining the chemical composition of the continental crust and its underlying section of mantle. □

Richard W. Carlson is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington, DC 20015, USA.

CYTOKINES

Can there be life without LIF?

John K. Heath

LEUKAEMIA inhibitory factor (LIF) is a polypeptide growth factor with a seemingly remarkable range of biological actions in different tissue systems. From studies in tissue culture these include (in rough developmental sequence) the pluripotential stem cells of the early embryo, primordial germ cells, peripheral neurons, osteoblasts, adipocytes, hepatocytes and endothelial cells, as well as various leukaemic cells of the myeloid lineage¹. Perhaps the most dramatic manifestations of the diverse action of LIF result from chronic administration of LIF to adult animals. These manifestations include rapid weight loss, behavioural disorders, ectopic calcification and bone abnormalities².

If too much LIF is clearly bad for you, what are the consequences of having no LIF at all? The answer, rather unexpectedly, is remarkably little in respect of the known targets of LIF action, but quite a lot in terms of female reproductive physiology. On page 76 of this issue, Stewart *et al.* report that female mice harbouring genetically inactivated LIF genes have a specific defect in uterine function which prevents implantation of the developing embryo³. This result not only raises questions about the biological activities of polyfunctional cytokines like LIF but should lead into a better understanding (and ultimate control) of uterine function at the time of implantation.

Stewart *et al.* generated inactivating mutations in the gene encoding LIF using the techniques of gene targeting in embryonic stem cells³. Breeding chimaeras transmitting the mutant gene led to viable homozygous animals lacking functional LIF protein. These animals appear normal (aside from a small decrease in body mass) with no obvious signs of the anticipated dire phenotypic consequences of LIF deficiency. But, although adult males are fertile, adult

females are infertile as a result of a failure of the embryo to implant. As the embryos themselves remain viable within the uterus, the phenomenon is somewhat reminiscent of the naturally occurring phenomenon of delayed implantation observed in many mammalian species. The phenotype arises as a consequence of a specific defect in the maternal host rather than in the embryo because homozygous mutant embryos implant and develop normally in a normal uterine environment, but wild-type embryos fail to implant in mutant mothers. Moreover, injection of LIF into female LIF-deficient mice at the period of implantation leads to partial rescue of the phenotype, in which embryo implantation and limited postimplantation development can then occur.

This phenotype is not totally unexpected if the normal pattern of LIF expression is taken into account. It has proved difficult to demonstrate convincingly that LIF is expressed in many of the tissues in which it is supposed to be active. The exception is the uterus, in which there is a dramatic, maternally controlled surge of LIF expression in the endometrial glands a few hours before implantation, which persists over the period of implantation before subsiding back to very low levels as embryonic development proceeds⁴. This burst of LIF is presumably required in some way to prepare the uterus for implantation, so without it implantation would be blocked. Although the physiology of the mouse uterus may be different from that of the human, it will be interesting to find out the extent to which LIF malfunction is involved in human reproductive failure and to see if implantation rates can be improved by administration of exogenous LIF over the implantation period.

A more general issue to arise from this

result is how a growth factor can have so many functions that become obvious from its overexpression, and yet so few when tested by loss of function. It is becoming clearer as more and more genes succumb to gene targeting techniques that control system design in multicellular vertebrates is impressively sophisticated. It may be that many intercellular control systems have a powerful compensation facility by which the loss of one participant is redressed through altered action of another factor with analogous functions. In the case of LIF, a number of understudies are waiting in the wings: in many cases the actions of LIF *in vitro* can be reproduced by one or more of a select group of growth factors, which includes interleukin-6 and interleukin-11, oncostatin M and ciliary neurotrophic factor.

This capacity to swap functions has recently acquired a molecular foundation with the appreciation that at least interleukin-6, oncostatin M, ciliary neurotrophic factor and LIF share a common feature in their action: they all use the transmembrane glycoprotein 'signal converter' molecule gp130 as part of a mechanism to transduce growth factor binding to the outside of the cell into intracellular molecular signals^{5,6}. The current data are consistent with a model for the action of this family of growth factors, in which association of the ligand with a 'private' ligand-specific receptor leads to the formation of a signalling complex including a 'public' class-specific transducer in the form of gp130.

In this model all members of a growth factor family sharing a common 'public' transducer can compensate for each other, providing the appropriate 'private' receptor chains are expressed. So whereas genetic inactivation of other members of the gp130 family of ligands might have specific, but relatively restricted, phenotypic consequences *in vitro* compared with their range of action *in vivo*, inactivation of gp130 would lead to widespread defects encompassing the combined cellular targets of the entire family. These models will doubtless shortly be tested experimentally, but in the meantime the LIF-less mouse reminds us that we can perhaps have too much of a good thing. □

John K. Heath is in the Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.

1. Polyfunctional Cytokines: IL-6 and LIF Ciba Foundry Symp. 167 (Wiley, 1992).
2. Metcalf, D. & Gearing, D. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5948–5952 (1989).
3. Stewart, C. L. *et al.* *Nature* **359**, 76–79 (1992).
4. Bhatt, H., Brunet, L. & Stewart, C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11408–11412 (1991).
5. Gearing, D. *et al.* *Science* **255**, 1434–1437 (1992).
6. Ip, N. *et al.* *Cell* **69**, 1121–1132 (1992).

Flashing giants forge fluorine

G. J. Mathews

FLUORINE is one of the few chemical elements whose stellar origin is still being debated. There are two reasons for this. For one, the only stable fluorine isotope, ^{19}F , is fragile and easily destroyed by proton capture reactions at stellar temperatures. Most elements of similar mass to fluorine are produced during the late stages of thermonuclear burning in massive stars. However, fluorine is so quickly destroyed that it tends to be bypassed by normal stellar evolution. The second reason is that, although clues to the origin of many elements can be obtained from stellar surface abundances, it has been difficult to find fluorine in stars, owing to a lack of useful atomic fluorine lines in the visible region of the spectrum. This latter difficulty was overcome recently by Jorissen, Smith and Lambert¹, who obtained fluorine abundances on the surfaces of various red giant stars from the infrared rotation-vibration lines of the HF molecule. Their measurements con-

stitute the only available information on the abundance of fluorine outside the Solar System, and they provide strong evidence that fluorine is produced deep within the interiors of red giant stars during episodic flashes of thermonuclear burning.

Red giant stars are identified according to their spectral class (or redness). In order of increasing redness (or decreasing surface temperature), the types of red giant stars studied in this work were those denoted K, Ba, M, MS, S, Sc, N and J. These different spectral types can roughly be identified with ages of red giant stars as they evolve along what is called the asymptotic giant branch² (see box). This is a phase in which stars become progressively more luminous and red as thermonuclear burning occurs in shells outside of an inert carbon-oxygen core while the outer envelopes become extended. There is a peculiarity in the structure of these stars^{2,3}. Helium burning (by which three helium nuclei

are fused into a carbon nucleus) occurs in such a thin shell that as energy is released, the temperature rises much faster than the star can expand and cool. This leads to a thermonuclear runaway which is arrested only by the development of a deep convection zone which dredges material from the thermonuclear burning regions all the way to the surface of the star. This process of episodic helium-burning shell flashes implies that as these stars evolve, their surfaces become enriched in the products of thermonuclear burning.

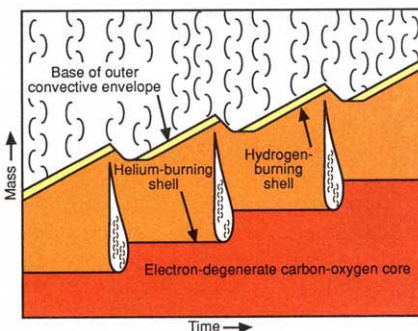
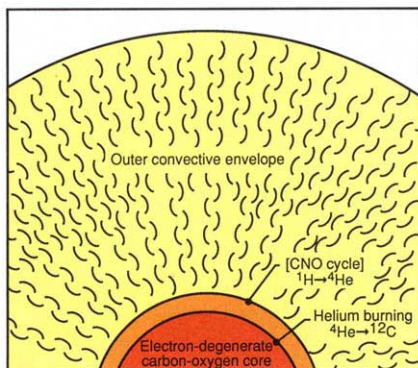
It is therefore significant that these measurements have revealed fluorine abundances as much as 30 times those in the Sun. This trend can be seen in the graph, taken from Jorissen *et al.*¹, which shows the logarithm of the ratio of fluorine to oxygen, $[\text{F}/\text{O}]$, (relative to the mean ratio for K and M red giants) as a function of the carbon to oxygen ratio, C/O . The various symbols represent different spectral types. For example, the open squares show K-stars whose fluorine abundance is not much different from that of the Sun. The more evolved M stars (open circles) show increases in fluorine and carbon abundances relative to oxygen as do S (closed circles) N (crosses) and J stars (open triangles). The trend of increasing carbon relative to oxygen with spectral type is an expected outcome of the dredge-up process which brings the carbon ashes of helium burning to the stellar surface. The correlation of increasing fluorine along with increasing carbon strongly suggests that fluorine is also being manufactured by the star.

An important question, however, is how fluorine could be produced and survive at the temperature of the helium-shell flashes. The authors have investigated several possibilities and conclude that the most likely reaction sequence involves the capture of an α -particle (helium nucleus) by ^{14}N to give radioactive ^{18}F , which decays to give ^{18}O ; proton capture by this then leads indirectly to ^{19}F — denoted $^{14}\text{N}(\alpha, \gamma)^{18}\text{F}(\beta^+)^{18}\text{O}(\text{p}, \alpha)^{15}\text{N}(\alpha, \gamma)^{19}\text{F}$. The protons are produced by a pair of reactions involving ^{13}C and ^{14}N . The fact that there are only a few protons present and that the circulation removes fluorine from the high-temperature regions saves the fluorine from destruction by proton capture. The fact that fluorine has not been destroyed by the $^{19}\text{F}(\alpha, \text{p})^{22}\text{Ne}$ reaction also implies a constraint on the temperature and neutron source reaction for the slow neutron-capture (s-process) nucleosynthesis of heavy elements⁴⁻⁶. At temperatures high enough to produce neutrons via $^{22}\text{Ne}(\alpha, \text{n})^{25}\text{Mg}$ in helium-shell flashes, fluorine would be destroyed. Thus, ^{22}Ne can not be the neutron source for the s-process. This is

Asymptotic-giant-branch stars

Top, structure of a red giant star as it evolves along the asymptotic giant branch. The outer convective envelope extends to hundreds of times the diameter of the Sun and continuously mixes material from its base to the surface. Below this, a hydrogen-burning shell produces helium (predominantly via the 'CNO' cycle for stars more massive than the Sun). The radius of this shell is actually only a few times the radius of the Earth and would barely be visible if this figure were drawn to scale. This radius, however, encompasses a mass comparable to that of the Sun. Slightly below the hydrogen-burning shell (a fraction of an Earth radius and a thousandth of a solar mass away) is a shell burning helium into carbon. All of this rests on an inert core of mostly carbon plus some oxygen formed during helium burning.

Below, evolution of the helium- and hydrogen-burning shells. Periodically (every 30–100 thousand years, depending upon the mass of the star) the helium shell ignites. Because the star can not expand fast enough to dissipate the heat generated, a thermonuclear runaway ensues. This is arrested by the development of a convective shell (teardrop shape in this figure). The hydrogen-burning shell then temporarily ceases and the outer convective zone briefly dips into the region below the hydrogen-burning shell. This process dredges up material from the helium-burning regions to the stellar surface. The observations of fluorine enhancement along with carbon on the surface of evolved red giants strongly suggests that fluorine is produced with carbon as a byproduct of helium burning and subsequently dredged to the surface.

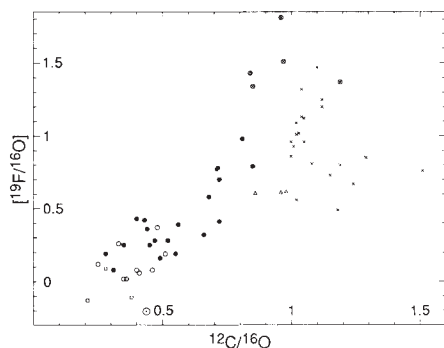


G.J.M.

SUPERANTIGENS

Playing upon both sides

Ursula Esser and Peter Parham



Logarithm of the $^{19}\text{F}/^{16}\text{O}$ ratio relative to the mean value in K and M stars plotted as a function of the $^{12}\text{C}/^{16}\text{O}$ ratio. The symbols are for stars of different spectral types. Open squares are K stars, closed squares are Ba stars, open circles are M stars, closed circles are S stars, circled crosses are SC stars, crosses are N stars, and open triangles are J stars. The trend of increased fluorine abundance with increasing carbon abundance and spectral type is strong evidence that fluorine is produced in helium-shell flashes within these stars.

consistent with other considerations⁴⁻⁶.

These data also suggest a problem for competing schemes for the formation of fluorine in stars. For example, it has been proposed by Woosley *et al.*⁷ that fluorine could be produced by neutrino-induced nucleon emission from ^{20}Ne during a type II supernova explosion. If that were true then the fluorine abundance should be correlated with other elements produced in supernovae (like oxygen)^{5,6}. One star in the sample was depleted in oxygen by roughly a factor of two. If fluorine were produced in supernovae, the fluorine to oxygen ratio should have been about the same as the Solar System value. Instead, the ratio is less, suggesting^{5,6} that fluorine was produced more slowly than oxygen. This points to the slowly evolving low-mass AGB stars as the source for fluorine rather than the rapidly evolving massive stars responsible for type II supernovae.

So, if anyone ever asks where the fluorine for your fluoride toothpaste comes from, you can answer that fluorine is probably formed in the frequent furious fusion flashes of an inflated flacid star, and on the new evidence you would probably be right. □

G. J. Mathews is in the Physics Department, Lawrence Livermore National Laboratory, Livermore, California 94550, USA.

- Jorissen, A., Smith, V. V. & Lambert, D. L. *Astr. Astrophys.* **261**, 164–187 (1992).
- Iben, I. & Renzini, A. A. *Rev. Astr. Astrophys.* **21**, 271–342 (1983).
- Schwarzschild, M. & Härm, R. *Astrophys. J.* **142**, 855–867 (1965).
- Käppler, F., Beer, H. & Wisshak, K. *Rep. Prog. Phys.* **52**, 945–1013 (1989).
- Mathews, G. J. & Cowan, J. J. *Nature* **345**, 491–494 (1990).
- Mathews, G. J., Bazan, G. & Cowan, J. J. *Astrophys. J.* **391**, 719–735 (1990).
- Woosley, S. E., Hartmann, D. H., Hoffman, R. D. & Haxton, W. C. *Astrophys. J.* **356**, 272–301 (1990).

In the wars between host and pathogen, whose side are superantigens on? In the case of the superantigens of mouse mammary tumour retroviruses, the answer seems to be 'both'. The picture emerging from recent reports^{1,2} is not only of how superantigens facilitate virus infectivity, replication and survival, but also of how they may be kidnapped and exploited by the host. More broadly, we can see a meeting of minds on superantigens, for the topic excited considerable debate at a conference dealing with the interplay of microbes and their hosts, held a month ago in Montana. After many years of going their own ways, microbiologists and immunologists are taking an interest in each other's wares.

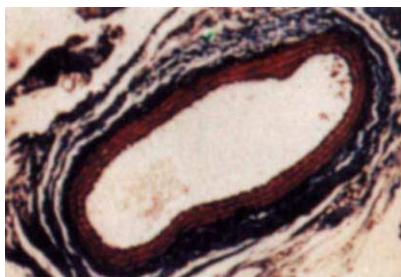
Superantigens bind to class II major histocompatibility complex (MHC) molecules stimulating families of T cells bearing particular V β chains. These potent proteins can derive from bacteria or viruses, and they differ from ordinary antigens by the greater numbers of T cells they trigger, their lack of processing into short peptides and the different sites of interaction with T-cell receptors and MHC molecules³. In humans, staphylococcal enterotoxins are bacterial superantigens active in food poisoning and toxic shock syndrome, while on the viral front, rabies⁴ and HIV^{5,6} are getting in

on the act. But, at the moment, it is on mouse mammary tumour viruses (MMTVs)^{7,8} that the spotlight has fallen.

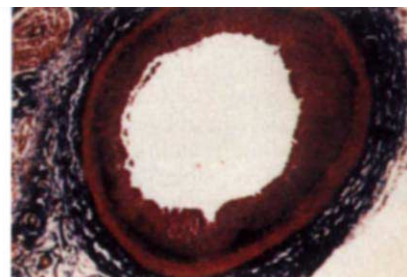
The MMTVs and their superantigens come in different guises; the exogenous viruses are inherited as infectious particles transmitted through mother's milk, whereas the endogenous viruses exist as proviruses integrated into the host genome and are passed on through its replication^{7,8}. T cells with antigen receptors using complementary V β chains are stimulated by a superantigen, and the consequence of such encounters depends upon the developmental stage at which they occur. 'Natural' exposure to bacterial superantigens often involves mature T cells and results in a strong polyclonal immune response³. In contrast, the superantigens of endogenous and exogenous MMTVs are encountered early in life and produce deletion or inactivation of the complementary T-cell families. In this manner the roster of MMTV carried by strains of mice — and there are many differences between them — shapes their T-cell repertoires, each virus determining the fate of part of the total T-cell pool.

Patterns of V β deletion can be manipulated experimentally by making mice transgenic for MMTVs. An extension of this approach has shown that the pre-

Anti-gene therapy?



Antisense



Control

ANTISENSE DNA has been used successfully to study the function of genes in intact cells, but its use as a therapy has been stymied by difficulties in delivery to the appropriate cells or organs — until now. On page 67 of this issue, M. Simons *et al.* describe how they have delivered antisense oligonucleotides to injured carotid arteries in rats. Injury such as that caused by balloon angioplasty, a therapy commonly used to 'widen' arteries in people, causes proliferation of the cells of the inner vessel wall, as is apparent in the right panel, with dangerous consequences. By painting an antisense oligodeoxynucleotide directed against the *c-myc* proto-oncogene onto the stripped and exposed artery wall in the injured region in a hardening, non-toxic gel, Simons *et al.* prevented the proliferation of cells in the vessel wall (left panel). After entering the cells, the antisense oligonucleotide is presumed to hybridize specifically to the messenger RNA encoding *c-myc*, thereby preventing its translation into protein. This approach therefore provides both a new route to cardiovascular therapy and a means to studying gene function in intact animals.

T. L. B.

vously mysterious⁸ open reading frame (*orf*) in the 3' long terminal repeat of these retroviruses is the source of the superantigen⁹. Results of other analyses¹⁰ pointed to this conclusion, but Acha-Orbea *et al.*⁹ have directly demonstrated that superantigen-mediated deletion is conferred upon mice made transgenic for the *orf*. The encoded protein is a membrane glycoprotein in which sequence variation in the carboxy-terminal region correlates with superantigenic specificity. So, viruses which delete different families of T cells differ in this region, whereas those affecting the same T cells tend to have similar sequences.

This rule, however, is not inviolate, as there is a case of dissimilar superantigens which delete the same T cells¹¹. Correlation of sequence variation in the carboxy-terminal 30 or so amino acids of the superantigen protein has raised the possibility that short, processed peptides derived from this region may, in the fashion of ordinary antigens, be the active moieties³. This model distinguishes the viral superantigens from their bacterial counterparts, for which it is clear that the native protein binds to class II MHC molecules outside the peptide-binding groove¹². The absence of experiments ascribing binding or superantigen activity to such peptides, synthetic or natural, combined with new results from Korman *et al.*¹, add weight to the view that the intact *orf* protein is the superantigen.

Korman *et al.* found that the *orf* protein has its C terminus and most of its mass on the extracellular (luminal) side of the cell membrane. So it seems poised to interact with the extracellular domains of the class II MHC molecule and engage the receptors of passing T cells. Interestingly, this relatively uncommon topology is similar to that of the invariant chain, an intracellular chaperonin and targeting protein, which also binds class II MHC molecules.

In suckling mice the port of entry for exogenous MMTV is the gut, whereas its subsequent site of replication is mammary gland epithelium. Although the precise mechanism of transport from gut to mammary gland remains obscure, involvement of the immune system, and T cells in particular, is apparent^{7,8}. One hypothesis is that T cells provide the transport. Infection of T cells by exogenous MMTV may be facilitated by activation, and in this manner the superantigen is postulated to be of benefit to the virus, and cost to the host.

Support for this view comes from Golovkina and colleagues². The authors characterized mice transgenic for the superantigen of an exogenous virus which deletes V β 14-bearing T cells. Varying levels of transgene expression correlated directly with the efficiency of

T-cell deletion and inversely with infectivity by the exogenous virus. These measurements demonstrate that the pathogenicity of an exogenous virus depends upon the presence of complementary T-cell receptors in the peripheral circulation. They also reveal how integration of MMTV as a provirus may be of selective benefit to the host in responding to the exogenous virus. Establishment of the provirus could break the viral life cycle by destroying its transport by T cells.

Instigation of mammary tumours by exogenous MMTV is considered to be a side effect of the viral life cycle, and one endowing little benefit to either virus or host. Prodigious quantities of virus are made by infected mammary epithelium and stochastic integration of virus near proto-oncogenes in this somatic tissue causes the tumours. Maybe occasional integration of virus into the host's germ line (by T-cell transport?) is another side effect, but one to succour the host—or more accurately her daughters. In the progression from infectious exogenous virus to uninfected provirus, the host subsumes the foreign as self. Viral superantigens essentially become self-superantigens and mice unable to beat the MMTV have apparently elected to join them. This evolution results in another side effect, the forfeit of T cells and modification of the immune system.

In MMTV infections T cells complementary to viral superantigens seem to play a beastly role, and from the results of Golovkina *et al.* one can be persuaded that their targeted deletion is a jolly good thing. But what are the consequences for the immune response to other infections? In truth the answers are likely to be mixed, for every T cell has the potential to serve both friend or foe — so, depending on the precise circumstance, their absence or presence will help host or microbe. □

Ursula Esser is in the Department of Microbiology and Immunology, and Peter Parham in the Departments of Cell Biology, and Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA.

1. Korman, A. J., Bourgarel, P., Meo, T. & Rieckhof, G. E. *EMBO J.* **11**, 1901–1905 (1992).
2. Golovkina, T. V., Chervonsky, A., Dudley, J. P. & Ross, S. R. *Cell* **69**, 637–645 (1992).
3. Herman, A., Kappler, J. W., Marrack, P. & Pullen, A. M. *A. Rev. Immun.* **9**, 745–772 (1991).
4. Lafon, M. *et al. Nature* **358**, 507–510 (1992).
5. Imberti, L., Sottini, A., Bettinardi, A., Puoti, M. & Primi, D. *Science* **254**, 860–862 (1991).
6. Laurence, J., Hodssev, A. S. & Posnett, D. N. *Nature* **358**, 255–259 (1992).
7. Acha-Orbea, H. & Palmer, E. *Immun. Today* **12**, 356–361 (1991).
8. Coffin, J. M. *Science* **255**, 411–413 (1992).
9. Acha-Orbea, H. *et al. Nature* **350**, 207–211 (1991).
10. Choi, Y., Kappler, J. W. & Marrack, P. *Nature* **350**, 203–207 (1991).
11. Gollob, K. J. & Palmer, E. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5138–5141 (1992).
12. Dellabona, P. *et al. Cell* **62**, 1115–1121 (1990).

Natural deception

COSMETIC breast enlargement, says Daedalus, is not a recent invention: nature got there first. In the endless evolutionary battle of the sexes, the female breast has acquired a loading of passive fat, purely to deceive men into overestimating its owner's lactation potential. Modern silicone gel implants merely amplify this deception; sadly they often leak or give trouble later. So Daedalus is returning to first principles. His new cosmetic implant material will go where nature intended: within the many empty or incompletely filled fat cells of the breast. The resulting enlargement will be entirely convincing.

The body's fat deposits equilibrate rapidly. A healthy polyunsaturated fat diet, for example, will rapidly load your adipose tissue with polyunsaturated fat. Daedalus points out that such fats, particularly where the double bonds are close to the carboxylic fatty-acid group, are somewhat light sensitive. Light can split off the acid group, leaving a hydrocarbon oil. It can also crosslink the molecules into waxy polymers. So DREADCO's chemists are devising a cunning 'Photofat', easily decarboxylated and polymerized by light. It will be incorporated into a variety of special foodstuffs: margarine, chocolate, cakes and so on, all heavily coloured to shield their fat content from the light. A regular consumer will soon accumulate quite a lot of Photofat in her adipose tissue.

The second component of the treatment is DREADCO's special internally luminous bra. Its gentle local radiance, at the optimal reactive wavelength, steadily illuminates the translucent fatty tissue beneath. The Photofat in each fat cell will be converted to inert, biologically invisible hydrocarbon oils and waxes. The cells will take up more fat to replace their apparently lost content; this will be photoconverted in turn. The filling, expanding cells will slowly and subtly transform the wearer's figure.

The trick should work on almost any part of the body. In a joint venture with a leading fashion house, DREADCO is marketing a range of chic black leotards and body stockings fitted with internally luminous panels. As long as the wearer maintains her Photofat-rich diet, the subcutaneous fat beneath the luminous regions will slowly expand. The areas under the black regions will be shielded from light, and will stay slim. Over several months, with no sudden changes to arouse suspicion, the wearer will build the figure of her dreams! It should then be almost permanent. The enlarged fat cells should retain their loading of inert, insoluble hydrocarbon oil for years. Any gradual loss could easily be topped up.

David Jones

Thyroid cancer after Chernobyl

SIR — We would like to report a great increase in the frequency of thyroid cancer in children in Belarus, which commenced in 1990 and continues. Table 1 shows the incidence of thyroid cancer in children in the six regions of Belarus and Minsk City from 1986 to the end of the first half of 1992. It can be seen that the overall incidence rose from an average of just four cases per year from 1986 to 1989 inclusive, to 55 in

problems, and is placing great strains upon the health services of our new country. It also provides an opportunity, which we hope will not be repeated, to study the consequences of major exposure of a population to isotopes of iodine from fallout. We are collaborating with several international groups and are preparing detailed reports of various aspects of the problem.

We believe that the only realistic

TABLE 1 Incidence of thyroid cancer in children in Belarus

Region of Belarus	1986	1987	1988	Years 1989	1990	1991	1992*	Total
Brest	0	0	1	1	6	5	5	18
Vitebsk	0	0	0	0	1	3	0	4
Gomel	1	2	1	2	14	38	13	71
Grodno	1	1	1	2	0	2	6	13
Minsk	0	1	1	1	1	4	4	12
Mogilev	0	0	0	0	2	1	1	4
Minsk City	0	0	1	0	5	2	1	9
Total	2	4	5	6	29	55	30	131

* Six months of 1992.

1991 and is projected to be not less than 60 in 1992. This increase is not uniformly distributed across the country: for example, there is no significant increase in Mogilev, Minsk City or Vitebsk. By far the greatest increase is seen in the Gomel region, from one or two cases per year to 38 in 1991, and a less obvious increase is seen in the Brest and Grodno regions.

The Gomel region lies immediately to the north of Chernobyl and is known to have received a high level of radioactivity as fallout after the breakdown of reactor number 4 on 26 April 1986. The plume passed first over the Gomel region in the first few hours after the major release of radioactivity, and then over the Brest and Grodno regions. The fallout contained large amounts of ^{131}I and significant amounts of the short-lived isotopes of iodine, although these were too short-lived to be measured.

We have classified the tumours according to the World Health Organisation classification (2nd edn) and find that virtually all are papillary carcinomas (128 of 131). They are, however, relatively aggressive, as can be seen from Table 2. Fifty-five of the 131 cases showed direct extension to the perithyroid tissues and six distant metastases, mostly in the lungs. It can be seen that only about 23 per cent were less than 1 cm in diameter. One of the children has died at seven years of age and ten others are seriously ill.

The occurrence of this increase in thyroid cancer in children within a few years of exposure to radioactive isotopes of iodine is unexpected, but real. It poses both humanitarian and scientific

explanation for the increase in the frequency of thyroid cancer is that it is a direct consequence of the accident at Chernobyl.

Vasili S. Kazakov

Ministry of Health of Belarus,
House of Government,
220010 Minsk, Belarus

Evgeni P. Demidchik

Thyroid Tumour Centre,
F. Skorinay Avenue 64,
220600 Minsk, Belarus

Larisa N. Astakhova

Radiation Medicine Institute,
Masherov Avenue 23,
220600 Minsk, Belarus

SIR — We have recently visited Belarus under the auspices of the WHO regional office for Europe and the Swiss government, and have had the opportunity to see some of the children with thyroid cancer, to study the pathology of the cases and to examine the relevant data.

We examined 11 children who had had operations for thyroid carcinoma and were now hospitalized for post-operative management or evaluation of metastatic disease. We were shown the complete records for these patients, including X-rays and echograms before and after treatment. All were diagnosed during the past 3 years, eight having been living in the Gomel region at the time of the Chernobyl accident and two in the Brest region. The age at diagnosis of the six females and five males was between 4 and 13 years of age; the youngest was born two days after the accident.

We have studied the histological slides from 104 cases of children from Belarus in whom the diagnosis of thyroid carcinoma had been made since January 1989. We agree both with the diagnosis of malignancy and of the type of malignancy in 102 of the cases. We also examined the data on the incidence of thyroid carcinoma in Belarus. There is a marked increase in frequency from 1990 onwards over the average for the years from 1986 to 1990. This increase started only 4 years after the Chernobyl accident, a surprisingly short time by comparison with studies of thyroid carcinoma that have followed exposure to external radiation in infants^{1,2}. Of the children with thyroid carcinoma in Belarus since 1990, the eight youngest at exposure were *in utero*, but were more than 3 months of fetal age at the time of Chernobyl. The fetal thyroid is known to start concentrating iodine at 12–14 weeks of gestation.

We do not believe that increased ascertainment of cases could have played more than a minor role in the recorded incidence of thyroid carcinoma. The proportion of resected nodules that are malignant is high and the type of tumour is aggressive. The ratio of thyroid carcinoma in children to that in adults has increased dramatically, although there are now signs that the incidence in patients over the age of 15 is beginning to increase. The rate is greatly in excess of the reported incidence of this disease

TABLE 2 Extent of spread (TNM classification) of thyroid cancer in children

	TNM symbol	Total number of cases	Lymph node metastases		
			None (N 0)	Ipsilateral (N 1a)	Other (N 1b)
Tumour size					
<1 cm	T1	30	17	10	3
1–4 cm	T2	33	17	8	8
>4 cm	T3	7	3	4	0
Extending to surrounding tissues	T4	55	14	18	23
Distant metastases	M1	6	1	1	4
Total		131	52	41	38

Classification as in *TNM Atlas* 3rd edn, eds Spiessl, B. et al., UICC (Springer, Berlin, 1990).

in children under 15 years of age, which is of the order of 1 per million per year³⁻⁶. In the Gomel region (total population about 2.5 million), the region of Belarus that received the highest fallout from Chernobyl, the incidence in 1991 and the first part of 1992 is approximately 80 per million children per year.

It is generally accepted that external radiation to the neck is associated with an increased incidence of thyroid carcinoma in man, and there is an increased sensitivity of the infant thyroid to the carcinogenic effect of radiation². In some animal studies, but not all^{7,8}, external radiation is found to be a more effective carcinogen for the thyroid than ¹³¹I. Clear evidence that the diagnostic or therapeutic use of radioiodine in man carries a carcinogenic risk is lacking^{9,10}, and ¹³¹I has provided a safe and effective treatment of Graves' disease in adults, although it is rarely used in young children.

The combination of the high level of exposure to radioactive fallout and the numbers exposed within a short time after its release makes the Chernobyl accident an unprecedented event. In the Marshall Islands, although the doses were probably comparable, the number of people exposed was several orders of magnitude smaller¹¹. In the case of the accident at Windscale (now called Sellafield), the number exposed was substantial but the doses were smaller¹², and no adequate study of any long-term thyroid effects has yet been reported. Other studies of fallout from weapons and of nuclear accidents (such as on Three Mile Island) have yielded inconclusive evidence. A close relationship between radiation dose and the incidence of thyroid carcinoma has been documented in atomic bomb survivors in Japan¹³, but the radiation received was mostly external and the contribution from fallout is uncertain.

We believe that the experience in Belarus suggests that the consequences to the human thyroid, especially in fetuses and young children, of the carcinogenic effects of radioactive fallout is much greater than previously thought. Studies of the Marshall Islanders, of the atomic bomb survivors and of the effects of external radiation on the thyroid suggest that the incidence of thyroid cancer in Belarus will be raised for many years.

The accident and its impact on Belarus poses a challenge to the international community to help, both in dealing with the extensive present and future public health consequences, and in promoting research for the understanding of the basic processes underlying the phenomenon. Understanding the consequences of Chernobyl will provide an important basis for preventive action in future.

Keith Baverstock

WHO European Centre for Environment and Health,

00156 Rome, Italy

Bruno Egloff

Pathology Institute, Kantonspital, 8401 Winterthur, Switzerland

Aldo Pinchera

Institute of Endocrinology, University of Pisa, 56100 Pisa, Italy

Charles Ruchti

Pathology Institute, University of Berne, 3010 Berne, Switzerland

Dillwyn Williams

Department of Pathology, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, UK

High-energy cosmic-ray origin

SIR — Although we have reasonable ideas about the origin of low-energy cosmic rays¹, there is considerable difficulty in determining their origin at the highest energies, above 10¹⁸ eV. Bryant *et al.*² have now suggested that these particles arise from a stochastic Fermi-type acceleration within our Galaxy, but we believe that there are serious problems with this hypothesis.

Bryant *et al.* treat the excess number of increasing energy scatterings, referred to as x , as a random variable instead of adopting its usual fixed average value³. They invoke a stochastic process whereby the acceleration can be unusually rapid for the particles in the high-energy wing of the statistical energy distribution, which they call 'high-flyers'.

Although it is true that the original theory can be improved in the sense of deriving a flatter energy spectrum, the

maximum energy is inviolate, being determined by the product of magnetic field and the linear dimension of the scattering centres. Unfortunately, and surprisingly, Bryant *et al.* choose the expectation value of x to be zero instead of βN_0 (βc is the velocity of the scattering centres and N_0 the mean number of collisions) and this causes a steeper spectrum, with index for the differential spectrum of $\gamma = 1 + \sqrt{1/(2N_0\beta^2)}$. The accurate value is, in fact,

$$\gamma = \frac{1}{2} \left(1 + \sqrt{1 + \frac{1}{N_0\beta^2}} \right)$$

In the event, Bryant *et al.* chose values for the parameters in their analysis which give rise to a spectrum identical in form to the standard Fermi value! Thus, the novelty of the approach is lost.

More seriously still, there are problems with the adopted parameters. To accelerate particles to energies of 10²⁰ eV or so, Bryant *et al.* take the following values for the scattering centres (or 'clouds'): radius $a = 10^{19}$ cm, mean spacing $d = 10^{20}$ cm, magnetic field $B_0 = 6 \times 10^{-2}$ gauss, velocity through the Galaxy $\beta c = 0.1c$. We argue that these values differ from the observational ones, dramatically, in the following ways.

(1) The mean energy density of the magnetic field averaged over the Galaxy would be 4×10^5 eV cm⁻³ for their more reasonable set of parameters, with $\beta = 10^{-4}$ (at the expense of a maximum energy of 2×10^{18} eV). These values are to be compared with ~ 1 eV cm⁻³ in gas motions in the interstellar medium, in starlight and in the cosmic rays themselves.

(2) The irregular component of the galactic magnetic field would be $B_{ir} \approx aB_0/d = 6,000$ μ G on a scale $L = 50$ pc, to be compared with 5 μ G inferred from the Faraday rotation measure of nearby pulsars^{4,5}, assuming a single-cell model for the field.

(3) The product of field times linear dimension, 180 mG pc, is very much greater than any known extended system in the Galaxy.

(4) In any reasonable model of the nature of the 'clouds', they will contain considerable masses of (ionized) gas — their consequent total mass and kinetic energy would be exorbitant. Finally, the clouds would be completely unstable.

We believe that second-order Fermi acceleration may work below 3×10^{17} eV in our Galaxy, an energy set by the Larmor radius of the particle trajectories, but that only a very small fraction of the particles are accelerated in this way. In fact, due to the well-known energy-dependent escape time, important contributions may be made only at energies below a much lower energy:

1. Shore, R. E. *et al.* *J. natn. Cancer Inst.* **74**, 1177–1184 (1985).
2. Ron, E. *et al.* *Radiat. Res.* **120**, 516–531 (1989).
3. Brown, P. D. *et al.* *Int. J. Epidemiol.* **18**, 546–555 (1989).
4. McWhirter, W. R. & Petroeschvsky, A. L. *Int. J. Cancer* **45**, 1002–1005 (1990).
5. Young, J. L., Ries, L. G., Silverberg, E., Horm, J. W. & Miller, R. W. *Cancer* **58**, 598–602 (1986).
6. Muir, C., Waterhouse, J., Mack, T., Powell, J. & Whelan, S. *IARC Sci. Publ.* no. 88, Vol. 5 (International Agency for Research on Cancer, Lyon, 1987).
7. National Council on Radiation Protection and Measurements *NCRP report no. 80* (Washington DC, 1985).
8. Lee, W., Chiacchierini, R. P., Shleien, B. & Telles, N. C. *Radiat. Res.* **92**, 307–319 (1982).
9. Holm, L. E., Dahlqvist, I., Israelsson, A. & Lundell, G. *New Engl. J. Med.* **303**, 188–191 (1980).
10. Holm, L. E. *et al.* *J. natn. Cancer Inst.* **80**, 1132–1138 (1988).
11. Conard, R. A. in *Radiation Carcinogenesis Epidemiology and Biological Significance*, Boice, J. D. & Fraumeni, J. F. eds (Raven, New York, 1984).
12. Baverstock, K. F. & Vennart, J. *Health Phys.* **30**, 339–344 (1976).
13. Ezaki, H., Ishimaru, T., Hayashi, Y. & Takeichi, N. *GANN Monogr. Cancer Res.* **32**, 129–142 (1986).

10^{11} eV. It seems to us that, at least below 10^{14} eV, first-order Fermi acceleration may dominate, this process occurring in supernova remnants.

In conclusion, second-order Fermi acceleration may work at low energies ($<10^{11} - 10^{14}$ eV) and contribute a small fraction of the Galaxy cosmic rays there, but it requires unphysical values for the parameters of the Galaxy at much higher energies. The problem of the origin of ultra-high-energy cosmic rays therefore remains.

X. Chi

C. Dahanayake*

J. Wdowczyk

A. W. Wolfendale

*Department of Physics,
University of Durham,
Durham DH1 3LE, UK*

* Permanent address: University of Kelanya, Sri Lanka.

SIR — Bryant *et al.*² have given an alternative treatment of Fermi acceleration based on an analogy between collisions of cosmic rays with clouds and coin tosses. We show elsewhere⁶ how this treatment may be extended to include the bias for head-on versus overtaking collisions that is an essential feature in Fermi's original treatment. When the bias is included, the result of Bryant *et al.* reduces to a one-dimensional version of a result that is well known⁷⁻⁹.

The suggestion by Bryant *et al.* (see also the accompanying News and Views by Zank¹⁰) that 'high flyers' in Fermi acceleration can account for cosmic rays up to $\sim 10^{20}$ eV is inconsistent with known results for Fermi acceleration. In the conventional stochastic treatment of Fermi acceleration, monoenergetic particles with initial energy E_0 experience a systematic acceleration and a diffusion in energy. After N collisions, systematic acceleration leads to a mean energy $E = E_0 \exp(\frac{1}{2}\beta^2 N)$, where the factor $\frac{1}{2}$ is a feature of the one-dimensional treatment, and the diffusion causes a spread with half-width $\Delta E \approx E\beta N^{1/2}$ about this mean. The 'high flyers' correspond to the diffusive high-energy tail.

Bryant *et al.* drew an incorrect conclusion about the effectiveness of the acceleration, by which they clearly meant the time for acceleration: by integrating over N , which corresponds to evaluating the asymptotic spectrum by integrating over time, they lost all information on how long it takes to set up the asymptotic spectrum. In fact, after a fixed number of collisions, the spectrum of 'high flyers' falls off exponentially, faster than any power law, and so necessarily faster than the asymptotic power law spectrum. The characteristic time for the asymptotic power-law spectrum to be set up at a given energy is the time required for systematic acceleration to that energy. The 'high flyers' can play no

significant role independent of the systematic acceleration.

Bryant *et al.* provide new insight into Fermi acceleration, but they report no new physics and cannot resolve the long-standing difficulties in the application of the Fermi mechanism to cosmic rays.

Lewis Ball

Don Melrose

Colin Norman*

Research Centre for Theoretical

Astrophysics,

University of Sydney,

New South Wales 2006, Australia

* Permanent address: Department of Physics and Astronomy, Johns Hopkins University and Space Telescope Science Institute.

BRYANT *ET AL.* REPLY — In our initially purely mathematical re-evaluation of the capabilities of Fermi acceleration², the key parameters were all derived, using essentially only the observed spectral slope; none was freely adopted, as appears to have been assumed by Chi *et al.* Preliminary results of a separate exercise to refine the *prima facie* conclusions in the light of astrophysical measurements are reported here.

We omitted the usual statistical bias favouring head-on, or energy-gaining collisions, because a preliminary Monte Carlo numerical study had shown it not to be essential to the formation of a power-law distribution. We readily find from a formulation⁶ including a systematic term that the value of $N_0\beta^2$ required to give the observed spectral slope is $\frac{1}{3}$, as opposed to $\frac{1}{2}$ with no bias. This increases E_{lim} , but only by 7 per cent. Deductions based on 'a fixed number of collisions' (to quote Ball *et al.*) seem to be invalidated by the finite, well-defined probabilities² for any number of collisions and for any combination of energy gains and energy losses. As the maximum number of collisions is limited, in principle, only by the age of the system, there seems to be no fundamental reason why the asymptotic power-law spectrum should not be reached.

Let us now consider the possible constraints imposed on the derived optimal or critical magnetic fields, H_{crit} , by observed Faraday rotation of pulsar radiation. For an average Faraday rotation to correspond to B_0 , space would need to be filled with magnetic fields of this magnitude. So, as the filling factor is $(\%a)^3$, the average Faraday rotation, and apparent field, will be reduced by this factor. Random orientations within a distance scale $l \gg d$ will lead to a further reduction, which from simple considerations may be estimated as $\approx 1/\sqrt{3} (\%a)^{3/2}$. The apparent field on the l scale is thus $B_l \approx k (\%a\sqrt{4\pi})^3 B_0$, where $k = 1/\sqrt{3}$ for $l \gg d$, and $k \rightarrow 1$ as $l \rightarrow d$. On the measurement scale of 50 pc ($1.5 \times$

10^{20} cm), the values of H_{crit} derived for $\beta = 10^{-1}$, 10^{-2} , 10^{-3} and 10^{-4} would appear as approximately 90, 6, 0.4 and 0.03 μG , respectively. (We have used here true values, rather than the rounded ones of our Table 1.) The last three are clearly permitted within the observational estimate of 5 μG (ref. 4). Further, the single-cell approximation used to estimate the field is considered to be too simplistic to give meaningful results, and the random field is actually more complex⁴, so even the $\beta = 10^{-1}$ model seems not too extreme. It may be prudent, though, for this case, to curtail the field strength by a factor of ten, and thus limit the maximum energy to 3×10^{19} eV. From the $\beta = 10^{-2}$ example, it is clear that proton energies up to 5×10^{19} eV are permitted by current knowledge of galactic magnetic fields. For heavy nuclei the limiting energy is higher, permitted values for iron, for example, extending up to $\sim 10^{21}$ eV.

To conclude, we respond individually to the numbered comments of Chi *et al.* (1) The energy densities of H_{crit} do, indeed, range from 3×10^4 eV cm^{-3} to 3×10^{-7} eV cm^{-3} , when the true (pre-rounding) filling factor of 3×10^{-3} is taken into account. If, as is implied, 1 eV cm^{-3} is taken as a criterion of acceptability, scattering-centre fields of ~ 100 μG would (with the above filling factor) be permitted. Taking up our earlier qualification² to the effect that, if the magnetic field strength is known to be less than H_{crit} , then the limiting energy is reduced in proportion, our four models yield maximum energies for protons of 4×10^{17} , 4×10^{16} , 4×10^{15} and 4×10^{14} eV. For iron nuclei these maxima become approximately 10^{19} , 10^{18} , 10^{17} and 10^{16} eV. These figures are encouragingly close to the highest observed values. (2) We have shown here that even the critical fields resulting automatically from the analysis (H_{crit}) are consistent with observation when the filling factor, $(\%a)^3$, rather than $\%a$, is used, and random orientation is allowed for. (3) The product of aH_{crit} is indeed, for the $\beta = 10^{-1}$ case, as quoted. But for fields of 100 μG , as already discussed, the product ranges from 3×10^{-4} mGpc to 0.3 mGpc for our four models. Such values seem to correspond, as far as we have been able to ascertain, to those of diffuse clouds and dark clouds. (4,5) The scattering centres referred to in our paper were not necessarily gas clouds, nor were they necessarily stable on cosmic timescales. It is possible that the inhomogeneities are, in fact, wavelike in nature, reflecting a degree of turbulence in the interstellar plasma. The scattering-centre treatment requires only that the inhomogeneities remain sensibly unchanged while being traversed by a cosmic ray. Stability timescales range,

therefore, from a few decades in the $\beta = 10^{-1}$ case to a few weeks when $\beta = 10^{-4}$.

D. A. Bryant

G. I. Powell

C. H. Perry

Rutherford Appleton Laboratory,
Chilton, Didcot, Oxon OX11 0QX, UK

1. Bhat, C. L., M. R., Mayer, C. J. & Wolfendale, A. W. *Nature* **314**, 515–517 (1985).
2. Bryant, D. A., Powell, G. I. & Perry, C. H. *Nature* **356**, 582–583 (1992).
3. Fermi, E. *Astrophys. J.* **119**, 1–6 (1954).
4. Rand, R. J. & Kulkarni, S. R. *Astrophys. J.* **343**, 760 (1989).
5. Chi, X. & Wolfendale, A. W. *J. Phys. G* **16**, 1409–1421 (1990).
6. Ball, L. T., Melrose, D. B., Norman, C. A. *Astrophys. J. Lett.* (submitted).
7. Davis, L. Jr *Phys. Rev.* **101**, 351–358 (1956).
8. Kardashev, N. S. *Sov. Astron.* **6**, 317–327 (1962).
9. Burn, B. J. *Astr. Astrophys.* **45**, 435–438 (1975).
10. Zank, G. *Nature* **356**, 564 (1992).

Parent–offspring conflict

SIR — Emlen and Wrege¹ describe a nice example of agreement between the behaviour of bee-eaters and parent–offspring conflict (POC) theory. But additional interpretations may be plausible. Of the 47 observed attempts at harassment, only 17 were by parents of the victims. Because this fraction is low, one might ask whether the situation described has originated and been maintained by selection directly for parental exploitation of offspring or by another route.

More apparent from the data is conflict between older and younger individuals, or generation conflict. Age alone explains more harassment (72%) than parent–offspring relatedness alone (36%). The fact that male offspring are sometimes harassed by their parents could be no more than a byproduct of harassment by breeders of younger individuals in their territorial group that attempt breeding. Harassers were said to have “preferentially selected close genetic kin as targets”, but this result seems to occur mainly from the delayed dispersal of sons and the fact that sons are necessarily younger than their fathers.

The benefit to a parent from a son who helps it after abandoning its own breeding effort would then be a simple consequence of delayed dispersal by sons, indiscriminate harassment of some younger breeders, and the tendency of sons to help parents. Because older individuals are assumed to be dominant over younger ones, an alternative hypothesis, dominant–subordinate conflict (DSC), is indistinguishable from generational conflict in this case. Harassed males were more likely to become helpers than unharassed males, but this observation is consistent with parent–offspring, generational and dominant–subordinate hypotheses, given the delayed dispersal of sons.

Because 24–31% of successful recruitment attempts reported in ref. 1 did not involve a parent and its offspring, it is clear that recruitment of non-offspring can be effective and that the generational and dominant–subordinate hypotheses would also ‘explain’ the behaviour of harassers by generating benefits.

By invoking selection theory, Emlen and Wrege have, perhaps unintentionally, implied that the intolerance shown by breeding bee-eaters towards sons has been increased and maintained by natural selection. In the avian context, however, the norm is for parents to exclude offspring totally from breeding on the parental territory. It follows, therefore, that the situation in bee-eaters constitutes a reduction in aggressiveness of fathers towards sons rather than an increase. A yearling paired male has a 74% chance of not having its nesting attempt disrupted (59% if parents present). Thus, although bee-eater parents do harass a few offspring, among others, this may be part of a general intolerance by dominants of subordinates who share their resources, and it may have arisen through increased tolerance of offspring rather than increased harassment of them, as has been suggested using parental-facilitation theory for New World jays². This theory accommodates both the facilitation of breeding by some offspring and the harassment of others.

Jerram L. Brown

Department of Biological Sciences,
State University of New York,
Albany, New York 12222, USA

EMLÉN AND WREGE REPLY — Virtually all cooperatively breeding species exhibit familial social organizations in which grown offspring delay dispersal and remain with their parents. This is generally considered to be the necessary precursor for the evolution of helping^{2–5}. Thus we agree that bee-eaters show increased parental ‘tolerance’ of offspring when contrasted with nonhelping species. But the question we addressed was not the origin of families, but the separate issue of the social dynamics of individuals living within families.

Cases of breeding harassment and suppression are commonplace within family-structured animal societies. Our goal¹ was to offer a framework for examining such familial conflicts and their resolutions based upon kinship. Brown offers an alternative framework based upon age and dominance.

Kinship and dominance undoubtedly have additive effects upon recruitment. Our data, however, refute the hypothesis that age/dominance is the primary influence, in that we reported¹ a significant tendency for harassing bee-eaters nonrandomly to target close kin as the subjects of their harassment. Perhaps we

did not state clearly enough that this analysis was restricted to harassment patterns within extended family groups and only to equal aged or younger family members. We have now repeated this analysis including only younger family members. In this test, a wide array of kin of differing degrees of relatedness are potentially available as targets, but all are younger than, and presumably subordinate to, their harassers. The results are the same, confirming that harassers are not “indiscriminate” in their interactions with younger family members, but preferentially harass those to whom they are most closely related.

The costs of ‘yielding’ to harassment explain this pattern. Targeted individuals that are unrelated will gain little from becoming helpers and are predicted to exhibit greater resistance (as demonstrated¹). Selection will then favour an ability of harassers to focus their efforts on those individuals likely to resist the least — namely close kin.

Parent–offspring conflict theory is merely a subset of inclusive fitness theory⁶. We did not mean to imply that father–son interactions have been moulded by selection pressures distinct from other categories of relatives. Our point was that kinship influences both the types of interactions expected with, and the amount of ‘leverage’ wielded over, others. Kinship-specific payoffs predict the pattern of harassment/recruitment found in bee-eaters. They also provide a testable framework for interpreting general patterns of reproductive conflict and suppression in all other animals that live in family-structured groups.

Stephen T. Emlen

Peter H. Wrege

Division of Biological Sciences,
Cornell University,
Ithaca, New York 14853, USA

1. Emlen, S. T. & Wrege, P. H. *Nature* **356**, 331–333 (1992).
2. Brown, J. L. *Helping and Communal Breeding in Birds* (Princeton University Press, 1987).
3. Stacey, P. B. & Koenig, W. D. (eds) *Cooperative Breeding in Birds: Long-Term Studies of Ecology and Behaviour* (Cambridge University Press, 1990).
4. Emlen, S. T. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 301–337 (Blackwell, Oxford, 1991).
5. Emlen, S. T. *Am. Nat.* **119**, 29–53 (1992).
6. Hamilton, W. D. *J. theor. Biol.* **7**, 1–52 (1964).

Scientific Correspondence

Scientific Correspondence is a relatively informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Because there is space to print only a small proportion of the letters received, priority is usually given according to general interest and topicality, to contributions of fewer than 500 words, and to contributions using simple language.

Genius with lucky chips?

Brian Oakley

Hard Drive: Bill Gates and the Making of the Microsoft Empire. By James Wallace and Jim Erickson. Wiley: 1992. Pp. 425. £14.95, \$25.95.

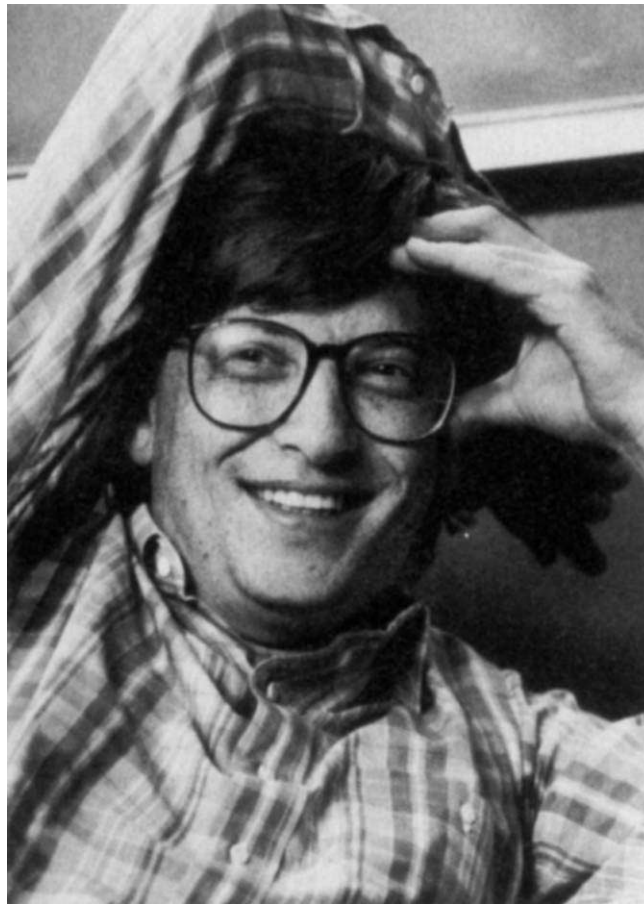
WHEN Napoleon Bonaparte said that his generals had to be lucky, he certainly intended them to manufacture much of that luck. Most people would say that Bill Gates, a self-made millionaire at 31, and now the fifth richest individual in the world at the age of 36, is among the lucky ones of the world. But was it luck and, if so, what created it?

In their fascinating book about Gates and his company, Microsoft, James Wallace and Jim Erickson try to answer this question. They claim that Gates is a technical genius. But the evidence is not impressive. True, he passed into Harvard University with a perfect score in his scholastic aptitude test, but perhaps the best clue to how he manufactures his luck comes from learning that he took the test twice because the first time he made a trivial error. Gates is not satisfied with anything less than a full score.

He, and his partner in the founding of Microsoft, Paul Allen, certainly had a clear eye for the future. As schoolboys they saw that the future of computing lay with the desktop computer, and before they left school were setting out to bring that about. Their first break came when they seized the opportunity to write a BASIC compiler for the hobby-kit microcomputer Altair that Ed Roberts and his company MITS were putting on the market in 1974. Typical of announcements about Microsoft products ever since, Gates and Allen told Roberts they had the compiler when, in fact, they had not written a line of the code. It took the two of them, working day and night in the Harvard computer centre, eight weeks to put together something that passed for a BASIC compiler for the Intel chip in the microcomputer.

Perhaps it did not require great genius to foresee the importance of the desktop machine. Although the authors do not refer to it, the idea had been around since 1945, when Vanessa Bell, then President Franklin D. Roosevelt's scientific adviser, wrote a remarkably prescient article predicting the advent of the

personal computer. By 1974, the idea was commonplace, with Alan Kay creating an early version, the FLEX machine, in 1969. Xerox had put together its great PARC team in the early 1970s, and it was this team that really provided the ideas behind many of the product innovations that made a fortune for



DOSSing about — Bill Gates in 1991.

Microsoft and for Gates.

Nevertheless, Gates and Allen did produce the BASIC compiler that fitted into a mere four-kilo words of memory, about 1,000 times smaller than the memory of most of today's computers (of which about 100 million now use Microsoft language compilers, Microsoft disk operating system (MS-DOS) and Microsoft user-interface Windows). It is doubtful whether anyone seriously used this BASIC compiler, but luckily for Gates and us all, the ever-growing rise in power of the microprocessor chip and the fall in memory costs ensured that subsequent versions could do a useful job. It would not be the only product

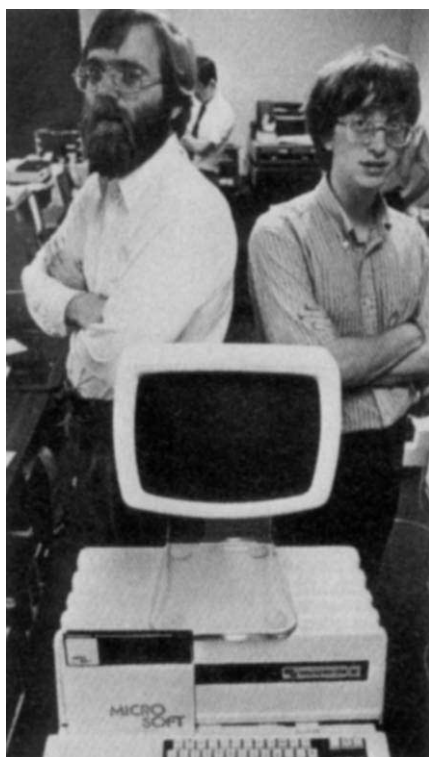
that Microsoft produced that was pretty poor to begin with, but which went on to dominate the industry.

Gates certainly foresaw the importance of the first IBM PC and, by a combination of hard work, enormous energy and drive — and perhaps a bit of luck in family connections — he was able to secure the contract to provide the operating system DOS. It was a cataclysmic and totally uncharacteristic error on IBM's part that it did not recognize that this operating system was to become a more important piece of property than the PC hardware itself. For most PC clones use Microsoft MS-DOS software, but IBM now supplies only a fraction of the hardware. Strangely, Gates himself did not recognize that the PC architecture was going to become standard, and went on writing software for other desktop machines that were doomed to fail. But perhaps he best showed his genius by grasping the importance of the Apple Macintosh computer, which used a revolutionary 'user-friendly' graphical interface — a concept that itself stemmed from Xerox's PARC. Microsoft still supplies much of the applications software for Apple, but Gates has effectively killed its key asset by developing his own 'windows' product that makes the PC readily usable. The latest version of Microsoft's Windows took six years to develop and required an investment of \$100 million. Gates has succeeded in creating *de facto* industry standards with DOS and Windows. It remains to be seen if he will succeed in his ambition to do the same for various application programmes, such as word processing and spreadsheet programmes. The jury is out, but judging by Microsoft's latest marketing ploy, he may well do so. If, like some 10

million other purchasers of the latest version of Windows, you recently received an offer to purchase Microsoft Word at almost half price, then you can see why the competition, such as the makers of WordPerfect, may feel that Microsoft's marketing ploys verge on the monopolistic.

It seems to be that enormous drive for dominance that forces Bill Gates on, rather than any great technical genius, or even a desire for money, for which he seems to have little personal use. Could a British software house have achieved the dominance that Microsoft has? Certainly the style of the company, as described by Wallace and Erickson, is not

Seattle Post-Intelligencer



Gates (right) with Paul Allen, co-founder of Microsoft, in 1981.

all that unlike a typical British software house of the same period. They make much of the easy interpersonal style, the free drinks, the large and boisterous parties. But these things happen in British software houses too. British software houses seem to be happier places in which to work, and they certainly produce good computer programs. Nor do they all lack vision. For example, Alex d'Agapeyeff at CAP steered the development of COBOL, or the common business-orientated language, until it became too expensive for his company. The work started about the time that Gates was launching Microsoft's BASIC compiler. Now there are said to be some 2,000 millionaires in Microsoft, and CAP has largely passed into French hands. The key difference seems to be that Gates always kept his eye on the market leaders, and in those days made sure he had a customer before he launched into a development, preferably a customer who could provide the necessary capital.

Wallace and Erickson's story makes captivating reading, more enthralling than any thriller. But it is only halfway through. We have seen the rocket rise; when will the stick fall back? What does a self-made billionaire in his thirties do for a new challenge? If Wallace and Erickson write the next chapter in the life of Bill Gates, I will certainly read it. □

Brian Oakley is at Logica, 68 Newman Street, London W1A 4SE, UK.

Dangerous democracies

John de la Mothe

Cardinal Choices: Presidential Science Advising from the Atomic Bomb to SDI.

By Gregg Herken. Oxford University Press: 1992. Pp. 317. £22.50, \$24.50.

It was in revenge for the theft of fire from heaven, we are told, that Zeus caused Hephaestus to make a woman out of Earth, who by her beauty should bring misery upon the human race. Hermes gave her boldness and cunning. Aphrodite gave her beauty and the gods called her Pandora — or 'all gifted'. When Epimetheus, brother of Prometheus, took her as his wife, he accepted from Zeus a wedding gift of a jar that contained every human ill. Alas, as is well known, the inquisitive Pandora opened the jar from which all miseries spread over the Earth. Only 'Hope' remained in the jar, which was shut quickly before she, too, escaped.

Time and again, this ancient myth has been used to describe the application of science and technology to the purposes of human contrivance. Of course, the 'guilt' of accepting Pandora's box lay strictly with the teacher, Epimetheus. In legend his personal fate is left unclear, but his dilemma remains with us still. Indeed, since antiquity it has been given to mankind to learn and thence to let our learning guide us in our fates. This ideology has been passed on through Newton, masked by Bacon and received today as a moral that should not be lost: those who love the beauty of their sciences must advise or teach us on the uses of the fire that their sciences contain.

But phrased in these terms — and no matter how attractive this myth has proved to be over the years in describing the 'ideal' relationship that 'should' exist between science and government — it is a myth that is, at best, instructive about morals and, at worst, totally uninformative in outlining political processes and structures that might help achieve these goals. And yet, at a time when industrial economies and democracies — both advanced and fledgling — need to encourage the development and diffusion of science, science advice to government needs most to be understood and improved, not mythologised.

But the mythologies that are derived from Prometheus are proving to be resilient. Following the famous development of the atom bomb, the 'balance of power' diplomacy of the Cold War in which "Mutually Assured Destruction" was the weapon of sane [sic] choice, and the political 'selection' of the "Strategic De-

fense Initiative", which seemed to some political 'masters and mandarins' to be the technical fix to beat all fixes, it is easy to see why historians and popular-science watchers would view such events as excellent illustrations of the process of science advising. Indeed, these are the three chosen case histories in Gregg Herken's new book on giving science advice to the US president. Unfortunately, they are the wrong cases.

There is no question that science advice to government has become of the utmost importance. The current debate in the United States, and within the nations of the Organisation for Economic Cooperation and Development, on international cooperation in big science, the migration away from military technologies and towards civilian technologies, the feared loss of technological leadership, technological trade balances and so on, all attest to the growing understanding within political circles of the socioeconomic importance of science and technology. Making the most of this understanding is now the game of science advising, but it is not an easy game.

Herken outlines as his framework the phrase from C. P. Snow's famous Harvard University Godkin Lectures, in which Snow said that the "cardinal choices" that governments must make are those that are the "sharpest and most dramatic"; they are those "that determine whether we live or die". There is no doubt that these words, when provided as a framework and coupled with a compelling set of historical vignettes, combine to promise readers a seductive and welcome analysis. But what is delivered is a light reading of Lord Snow, an abbreviated analysis of key events in history and a blind eye to how science advising works in most countries today.

Snow did indeed discuss the 'life and death' cardinal choices that government must make, but he also analysed the mechanisms through which decisions are taken. As he foresaw in 1960, the relationship between science and government was — and would increasingly become — a system of closed politics. This system can be typified in terms of types of power: committee politics, hierarchical politics and court politics. Of committee politics, he wrote that "the archetype of all these is the kind of committee where each member speaks with his individual voice, depends on his personality alone for influence, and in the long run votes with an equal voice." Of hierarchical politics, he said simply that this was the politics of the chain of command, of bureaucracy and of large industry. And he defined court politics as "attempts to exert power through men who possess a concentration of power". In reality, these three forms of closed politics interweave, interact and flow

from one to the other, regardless of objectives. All three form the functioning nexus of the science-government relationship. There is nothing mythical about this. The world works through decisions and this description simply reflects the way in which people have to operate in order to get anything done at all. This is, as Snow said, "how the world ticks".

As to the three principal cases in Herken's analysis, their most interesting aspects are clearly the awkward attempts of Leo Szilard, Albert Einstein, Niels Bohr, J. Robert Oppenheimer, Edward Teller and many others to affect the policy, politics and political decisions of the US government. There are many wonderful synopses of the scientists' feeling that "we could not assume responsibility", that they were "swimming in syrup", that they were "all green" and that "we did not even speak the same language". And in many ways, the cases chosen are superb examples of Snow's closed political system. But one cannot escape the sense that these cases are but synopses of important events, each of which already has a good number of more detailed and fascinating books to its credit. Nor can one escape the sense that little is made of the great potential to analyse the book's chosen cases in Snow's full framework. As Snow made clear, a nation with a concentration of personal power on matters of science has on its hands essentially a dangerous democracy. But although many of Herken's cases feature personalities with exactly such power, he ventures no such judgements about political validity.

Much has been written in recent years on the role of chief science advisers, on the proper structure of science advice and on the dissemination of balanced scientific information to all levels of government. In practice, the US president's power is countered by a system of checks and balances between Congress and the public. And in practice, there are hundreds of science-based departments and agencies in the United States, each of which filters and jockeys its views into the 'corridors of power' using mechanisms that are clearly 'Snovian'. And in practice, a great many scientists, science-policy analysts and policy advisers combine to give summary advice to the highest levels of government. Unfortunately, though promising, *Cardinal Choices* provides one more trip into Promethean territory, and not an illumination of science advice in modern times. □

John de la Mothe is in the Public Policy and Management Group, Faculty of Administration, University of Ottawa, Ottawa, Canada K1N 6N5 and at the J. F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA.



A sugar mill, from J. B. Dutertre's *Histoire générale des Antilles* (1667). The picture is taken from *Colonialism and Science* by James E. McClellan III, which explores the relationship between science and colonialism in the eighteenth-century French West Indian colony of Saint Domingue (modern-day Haiti). Published by The Johns Hopkins University Press, price \$52, £37.50.

Sex, soma and senescence

Darryl T. Gwynne

The Evolution of Life Histories. By Stephen C. Stearns. *Oxford University Press*: 1992. Pp. 249. £16.95, \$29.95.

I HAVE often thought that fields dealing with trait adaptation, such as physiological or behavioural ecology, could have more informative titles. 'Ecology' tends to direct one's thinking towards influences from the external (physical and biotic) environment. However, as we learn from this book, the most important aspect of the environment of any of an organism's traits are its other traits. Stearns studies evolution by examining adaptive compromises between individual life-history traits such as age, size and life span. In his view, this 'one trait at a time' analysis is partly a response to the failure of the approach in which all life-history traits are considered together (such as r and K selection theory).

Although aimed at advanced students, this work will be read much more widely, because the basics of life-history theory pervade all pure and applied branches of ecology, ethology and evolution. Stearns sets us on the right track for understanding interactions between traits, just as G. C. Williams did to trait selection in general with his *Adaptation and Natural Selection* (1966). But Stearns' review is not just a microevolutionary one. He weaves in an important chapter on phylogenetics and compara-

tive analysis, illustrating how past evolution can constrain continuing trait interaction.

It is not only the theme that makes the book an excellent read. I enjoyed the questions and comments designed to stimulate research and the occasional tales of, for example, food-unlimited kestrels and 200-year-old molluscs. The text flows well from start to finish because of clear summaries in, and connections between, the chapters. This is especially true of the first section on the fundamentals, which contains a lucid discussion of concepts such as natural selection, adaptation and fitness.

The reader is then introduced to trait evolution by learning how demography allows calculation of the intensity of natural selection and how the use of quantitative genetics can determine the response to this selection. Here, Stearns places genetics in nature by showing how genetic variation is expressed in different environments (using reaction norms to analyse plasticity). Compared with other (for example, metric) traits, life-history (fitness) traits are predicted to have reduced genetic variation. In reviewing the evidence, Stearns concludes that, although these traits do indeed have lower heritabilities, they can evolve because there is sufficient underlying genetic variation. But this conclusion is reached without any reference to comparative data on genetic variation. Heritability and genetic variation are not the same and, as it now turns out, lower heritabilities of fitness traits are

explained not by lower additive genetic variation but by the other sources of variation contributing to trait heritability. In the final analysis, fitness traits actually have higher additive genetic variation than other traits.

The essence of life-history study comes in chapter 4 and involves an understanding of the optimal allocation of resources in two or more traits. Studies of trade-offs use mainly genetic-correlation experiments or phenotype manipulations. The former method has been viewed as the more appropriate. But Stearns clearly considers phenotype manipulations as having a vital role and cites studies such as clutch-size experiments as outstanding accomplishments in life-history research.

Stearns admits that he excludes important topics, particularly modular organisms (such as clonal invertebrates) and complex life cycles. It seems fair, however, to view these topics as a 'next step up' from the solid foundation that this book builds. □

Darryl T. Gwynne is in the Department of Zoology, Erindale College, University of Toronto, Mississauga, Ontario L5L 1C6, Canada.

Multiple minds

Ian J. Deary

Intelligence and Development: A Cognitive Theory. By Mike Anderson. Blackwell: 1992. Pp. 256. £40, \$44.95 (hbk); £11.95, \$19.95 (pbk).

In his presidential address to the American Psychological Association in 1957, Lee Cronbach called for psychology's two scientific disciplines to recognize their different strengths and their need to integrate. The integration has been a long time coming; the relationship between differential and cognitive (née experimental) psychology is still characterized by much mutual ignorance and occasional antagonism.

Mike Anderson is an integrationist, and he has written a rich, multi-dimensional book about human intelligence. Material from differential psychology, developmental psychology, psychobiology and neuropsychology is used to build an original theory of the structure and development of human intelligence. He writes with a strong, clear voice and often addresses the reader directly, in a style reminiscent of Freud's *Introductory Lectures*. The reader is warned when things are likely to become technical, at times advised to skip sections if already happy to accept a proposition, and steered away from

metaphors that might seem too fanciful. (Intelligence attracts probably more metaphors than any other psychological phenomenon, with memory as a likely runner-up; nevertheless, Anderson's diverting account of intelligence as a law firm was new to me.) Adopting such a style must have been a calculated risk, inviting charges of flippancy. In fact, rigour of argument and quality of presentation do not suffer and, whereas some will find the style intrusive and annoying, most will find it engaging.

Evidence is gathered to support the following propositions about human cognitive abilities: cognitive abilities increase with development; individual differences are stable during development; cognitive abilities co-vary; there are specific cognitive abilities; and there are cognitive mechanisms that are universal for humans and which show no individual differences. All but the last of these is established satisfactorily and, black box by black box, a "minimal cognitive architecture" is sketched to link them. The theoretical centrepiece is a box labelled "knowledge". Apparently, there are two routes to it. First, knowledge is generated by running one of a number (which might be two) of "specific processors" on a "basic processing mechanism". Second, knowledge may be given directly by a number of dedicated processors that typically contain information that has been important during evolution. In contrast to the more customary rectangles that fence in the model's other components, these so-called "modules" are drawn with rather organic, potato-shaped outlines and contain labels such as "perception of three-dimensional space", "phonological encoding", "syntactic parsing" and "theory of mind".

The theory has merit, without necessarily being true. It brings together diverse information about human mental abilities and provides a clear organization. In addition, the theory generates at least one interesting hypothesis: that individuals with higher levels of general intelligence will show more cognitive differentiation. Some parts of the theory lack adequate evidence. Anderson says that there are no individual differences in the modules, yet this is never established, and it seems unlikely that there are no individual differences in, say, syntactic parsing. The modules and the specific processors are posited largely because of the existence of individuals with specific cognitive deficits and of idiots savant. Later, by tautology, Anderson argues that damage to the inferred structures explains these disorders. At times the theory is too fluid and imprecise, and leans too heavily on weak evidence. For instance, the cause of the basic processing mechanism's in-

tersubject variance seems sometimes to be speed, at other times efficiency and at still others capacity. In a daring prediction, its speed is hypothesized not to change during cognitive maturation; but to establish this, Anderson has to explain away good evidence to the contrary, and the only positive evidence would seem to be a single study where the key datum was a nonsignificant difference between two weak correlations.

The basic processing mechanism, the main source of individual differences in the theory, is no more than a reification of Spearman's *g*: "The basic processing mechanism represents a knowledge-free biological constraint on thought, and is responsible for the phenomenon psychometricians know as general intelligence." It is progress, however, to see a cognitive psychologist frame this psychometric discovery in a prominent box in the theory. The specific processors too, although their mechanism of interaction with the basic processing mechanism is novel, are no more or less than spatial and verbal ability, the two long-recognized prominent group factors in human ability. Here, as in other parts of the theory, the hard work is done by neuropsychology and differential psychology, which provide the evidence for the components, yet cognitive psychology gets the credit. The specific processors are brought in on the back of factor-analytical and clinical neuropsychology studies, yet they are said to be 'computational' of a nature that, although unknown, is not isomorphic with the types of ability that index them.

This is a symptom of the book's most poignant aspect. Anderson states at one point that the theory is "avowedly cognitive", yet later he has to confess that "[a]lthough computational in spirit, the theory does not, as yet, embody a computational model. That is to say, the theory could not be used at the moment to explain precisely how someone solves even the simplest of problems." Such a comprehensive failure of cognitive psychology to contribute anything much to this 'cognitive' theory is presaged earlier in the book when the reader is informed that "it would be a brave cognitive psychologist who would claim to have worked out, even in the broadest terms, an uncontentious processing theory of any significant cognitive ability". But this doesn't really matter. Anderson's faith that a cognitive account of his various components will eventually turn up does not prevent his doing a good job with what is to hand. And, who knows, his theory might even prove to be correct. □

Ian J. Deary is in the Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.

Irregular oscillations of the West Antarctic ice sheet

Douglas R. MacAyeal

Department of Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, Illinois 60637, USA

Model simulations of the West Antarctic ice sheet suggest that sporadic, perhaps chaotic, collapse (complete mobilization) of the ice sheet occurred throughout the past one million years. The irregular behaviour is due to the slow equilibration time of the distribution of basal till, which lubricates ice-sheet motion. This nonlinear response means that predictions of future collapse of the ice sheet in response to global warming must take into account its past history, and in particular whether the present basal till distribution predisposes the ice sheet towards rapid change.

CONCERN that the West Antarctic ice sheet (WAIS) might collapse in response to future climate warming is motivated by geological evidence suggesting its collapse some time during the past million years. (Here I use the term collapse to indicate the conversion by ice flow of a partially grounded, partially floating ice sheet into a purely floating ice shelf in such a manner as to increase the global ocean volume.) The evidence traditionally cited for possible WAIS collapse involves the interpretation of fossil beach and coral-reef terraces. Some interpretations suggest that sea level was 6 m higher during the 120 kyr BP (before present) interglacial (Sangamon, or isotopic stage 5e) than it is today¹. If the Greenland and East Antarctic ice sheets were roughly the same size during the Sangamon as they are today (an assertion subject to question²), such a high sea-level stand would point to a collapse of the WAIS³. This interpretation may be in doubt because the 6 m sea-level high-stand could actually represent a geodynamic effect associated with glacial rebound⁴ rather than a change in ocean volume. Nevertheless, support for possible WAIS collapse comes from the West Antarctic region itself: U/Th chronology of carbonate lacustrine deposits in the dry valleys⁵ (along the western margin of WAIS) confirms that there was a marked retreat of grounded ice between 130 and 98 kyr BP. Moreover, marine microfossils of age less than one million years are found in glacial till below grounded portions of the WAIS⁶. This suggests, but may not establish (L. H. Burckle, personal communication), seawater invasion at least once during the past million years.

Given concern over the future of the WAIS, and the likelihood that the geological record alone will not reveal the exact timing or cause of ice-sheet collapse, I use a numerical model of ice-sheet, ice-stream and ice-shelf dynamics to determine what attributes of dynamics and climate forcing could have caused WAIS collapse during the Late Pleistocene. I shall demonstrate that the long-term climate history of Antarctica associated with the glacial and interglacial cycle could have caused the WAIS to collapse and re-form several times over the past million years. The climate history is revealed by the surface temperature record derived from the Vostok ice core⁷, and the history of global sea level is estimated from isotopic records in deep-sea sediments⁸ (Fig. 1). The special characteristics of ice streams, most notably the dynamics introduced by a deformable bed, suggest a geological history that can include sporadic, perhaps chaotic, cycles of collapse and re-formation. The reason for this behaviour is that the timescale needed to equilibrate the distribution of deformable subglacial till is comparable to the 100,000-year timescale of the glacial cycle. The ice sheet thus cannot respond quasi-statically to long-term external climate forcing. Instead, it is jarred into possibly chaotic behaviour by the external forcing, and collapses occasionally when deformable basal till becomes widespread.

A whole ice-sheet model

To investigate the history of WAIS during the past million years, I constructed a finite-element model of ice-sheet flow and mass balance that can reproduce the present-day flow regime of the ice sheet, including its ice streams and ice shelves. This model represents a set of conservation rules which determine the time evolution of ice thickness in response to snow accumulation, basal melting, iceberg discharge, the horizontal divergence of ice flow dictated by gravitational driving stresses, and the temperature of the ice and its bed. Also included are conservation rules regarding the thickness and thermodynamic state of

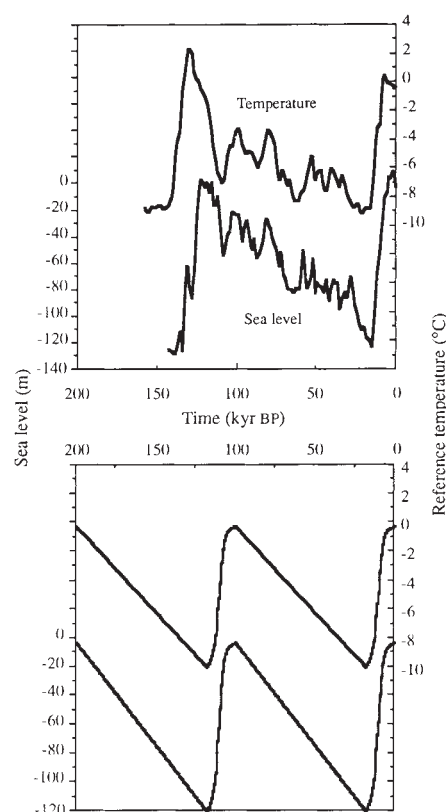


FIG. 1 Upper panel, Antarctic surface-temperature change chronology (upper curve) derived from the Vostok ice core⁷ and a global sea-level chronology (lower curve) estimated from deep-sea sediment isotope records⁸. Lower panel, abstract chronologies of surface temperature change (upper curve) and global sea level (lower curve) for the simple sawtoothed climate cycles used as forcing in the model simulation discussed in Fig. 3.

deformable subglacial sediments, which are the hallmark of WAIS ice streams⁹.

Inland-ice dynamics similar to those used successfully in previous studies of Antarctic glaciation¹⁰ determine the ice flow and mass balance where the ice sheet is frozen to its bed (basal sliding is not allowed). Ice-stream dynamics, including a treatment of the mechanics and thermodynamics of deformable subglacial till, are used where the ice-sheet bed reaches the melting temperature, and in elements that bridge the transition zones between wet and frozen bed¹¹. Ice-shelf dynamics are used where the ice sheet floats free of the subglacial bed¹². A heat-flow equation accounting for vertical conduction and advection predicts the temperature-depth profile in the ice. In the upper 2,000 m of the subglacial bed, this profile is modelled using a vertical heat conduction equation and a geothermal flux basal boundary condition (0.055 W m^{-2})¹³ to account for long-term thermal inertia effects. Horizontal heat diffusion and advection are disregarded because they cannot be treated accurately at the spatial resolution of this study. Latent heat and strain-heating effects in subglacial till are accounted for when determining the melting rate at the base of the ice. Crustal deformation in response to changing ice and seawater loads is treated in a simplified manner using an e-folding relaxation time constant of 8,000 years¹⁴. Elastic flexure and spatial length-scale effects¹⁵ are disregarded for simplicity.

The processes represented by this model are simplifications of those required for more comprehensive treatments of ice flow in West Antarctica. In my opinion, a simpler model could not capture the essential behaviour of the WAIS and a more complex model would be computationally unwieldy for million-year-long simulations (taking 25-year time steps). Work in which the model, or pieces of the model, has explained observations of the present WAIS^{10,11,16,17} defends my simplifications.

A critical feature of the model, and one in which it differs from previous models used to investigate WAIS, is its treatment of deformable basal sediments, the lubricant of ice streams. Relatively little is known about the mechanical and hydrological properties of deforming till layers beneath West Antarctic ice streams. Field observations suggest that they are ~5–10 m thick, are dilated and filled with pore water near the overburden pressure, and are so inviscid compared with ice that most of the horizontal ice velocity is accommodated by shear within the till^{18,19,26}. I account for four of the most important qualitative aspects of deformable bed physics. First, deforming till satisfies a mass-continuity equation (in which vertically integrated horizontal till-flux divergence is balanced by till-layer thinning and the creation or removal of deforming till at the till/rigid-sediment interface below the deforming till layer). Second, water is required to fill void space in newly dilated till (or is expelled when till ceases to deform and begins to consolidate). Third, strain heating is localized in till, and latent heat (of water in the pore space) is stored in till. Finally, till affects the basal friction of the ice sheet.

I use a vertically averaged continuity equation that treats horizontal till-flux divergence and erosion of the substratum to predict till-layer thickness. To compute till-flux, the vertically averaged horizontal velocity is assumed to be half the ice velocity at the ice/till interface²⁰. This assumption is correct if the till is a viscous fluid. For other rheological properties¹⁹, the vertically averaged till velocity would be less than the value assumed here. The conversion of rigid, consolidated substratum into the mixture of unconsolidated sediment and water that comprises deformable till is assumed to be limited by the availability of water. A rationale for this assumption is that water must be drawn through the relatively impermeable till to the till/rigid-sediment interface where void space is created. The pore-water pressure gradient necessary to induce this water flux exerts an important influence on the conversion rate²¹. The conversion rate is thus expected to be a function of the local basal-melting rate and the rate at which basal meltwater is dissipated by

subglacial drainage mechanisms. I assume a linear function. (More advanced models would require explicit treatments of the basal water conduit system, the pore-water pressure field, the basal stress imposed by the ice sheet, and the tenacity of the substratum.) If the basal melting rate is positive, the constant of proportionality is equal to $(1-f)/\theta$, where f is the fraction of total basal meltwater production dissipated by subglacial drainage and θ is the porosity difference between rigid substratum and till (θ is expressed as a non-dimensional ratio of pore volume to solid volume). If the basal melting rate is negative (freezing), I assume that subglacial drainage is shut down and the constant of proportionality is thus $1/\theta$. The value of f is not well constrained by observations, but the porosity of deforming till below ice stream B is 0.4 (ref. 22). To yield computed till thickness which falls in the observed range of 5–10 m, I take f and θ to be 0.75 and 0.4, respectively.

The effect of deformable basal till on the stress balance of the ice sheet is uncertain because of widely different interpretations of available data. I adopt the 'sticky-spot' interpretation, which holds that subglacial till is so inviscid that the large-scale basal friction is determined by isolated bedrock pinnacles that poke through the till layer, or by other lithological features of the glacial substratum¹⁹. Observations of basal-stress distributions¹⁷ suggest that sticky spots have small spatial scales (1–10 km) which fall below the finite-element resolution of the model used here. I therefore use a spatially uniform, empirically determined¹⁷ basal-friction coefficient ($1.5 \times 10^9 \text{ Pa s m}^{-1}$) in the ice-stream stress-balance calculation to represent the average large-scale effect of sticky spots. I define the basal-friction coefficient as a scalar that, when multiplied by basal ice velocity, gives the direction and strength of the basal traction. Although it is premature to accept the sticky-spot interpretation over other competing theories, I use it because it is simple, has been successful in simulating the observed velocity field of ice stream E (ref. 17) and gives the WAIS model the ability to produce till in amounts currently observed and to produce ice streams that move with the correct velocity scale. Sensitivity tests not reported here suggest that plausible alternative basal-till rheological

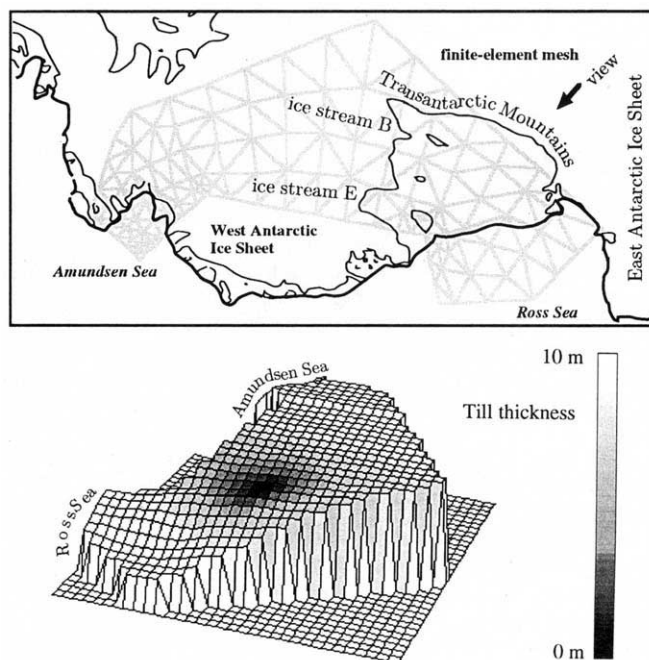


FIG. 2 Upper panel, west Antarctica and its low-resolution finite-element representation. Lower panel, an oblique view (taken from a viewpoint west of the Transantarctic Mountains) of the WAIS model surface elevation during typical interglacial conditions. Shading on the surface of the ice sheet represents the thickness of deformable subglacial till which lubricates the bed. Maximum till thickness in this particular configuration of WAIS (~10 m) occurs in roughly the area in which ice streams are found today.

properties, such as those of a purely viscous fluid²⁰, produce ice-sheet behaviour that is even more unstable than that described below.

Having simplified its physical processes, I also reduce the complexity of WAIS geometry to make the computation of WAIS evolution over long time periods manageable. Accordingly, I take the curved channel extending from the open Ross Sea to the Amundsen Sea through the region of present ice-stream activity (Fig. 2) to be the computational domain of WAIS. Most notably absent in the simplified geometry are the narrow ice streams which drain between interstream ridges along the boundary between inland ice and ice shelf. I forego the geometric but not the dynamical accuracy of these ice streams. As stated above, regions determined to rest on a melted bed are allowed to evolve according to the stress- and mass-balance equations that have successfully simulated ice streams in previous high-resolution models^{11,17}. Two boundaries of the idealized WAIS model domain separate WAIS from East Antarctica and from the relatively small but stable ice dome that sits on the coast between the Amundsen and Ross Seas. These boundaries are assumed to be rigid and impermeable; ice flow parallel and perpendicular to these walls is therefore zero in the model. The two boundaries that face the Ross and Amundsen Seas are treated as open, so that they freely allow the calving of icebergs into the ocean. For the stress balance at the open boundaries, I specify seawater pressure as the dynamic boundary condition. The undisturbed elevation of the continental shelf below the idealized WAIS geometry is assumed spatially uniform at a value of 350 m below sea level when ice is absent, isostatic rebound is complete, and global sea level is at its present value.

Climatic forcing is represented by two agents: local surface temperature (which affects local precipitation) and global sea level (which affects ice-sheet grounding lines). Following previous ice-sheet modelling conventions, ice-sheet surface temperature is taken as a function of ice-sheet elevation, an atmospheric lapse rate, and a reference sea-level air temperature which is assumed to vary in time according to the Vostok ice-core record of temperature change (Fig. 1). I prescribe a standard lapse rate for atmospheric temperature based on empirical data

($5^{\circ}\text{C km}^{-1}$ below 1,500 m and $14^{\circ}\text{C km}^{-1}$ above)²². This lapse rate has become a standard attribute of local Antarctic climate which allows intercomparison of Antarctic ice-sheet modelling studies. The reference sea-level air temperature (-22.5°C) is assumed to be spatially uniform, but is allowed to vary in time by up to 12°C as suggested by the Vostok ice-core temperature record⁷. Following previous ice-sheet modelling conventions, surface accumulation is specified to be the sum of a spatially uniform value of 0.25 m yr^{-1} (ice equivalent) and deviation which is a function of the local air temperature at the ice-sheet surface^{7,13}. As with the lapse rate, this deviation function represents a standard attribute that allows model intercomparison.

Global sea-level forcing is prescribed by changing the elevation of the continental shelf on which the ice-sheet rests. This forcing is assumed to be caused by the advance and retreat of the Northern Hemisphere ice sheets. Effects of the WAIS on global sea level are disregarded because they are small (5% of the typical glacial to interglacial sea-level change).

Response to 100-kyr climate cycles

To determine what attributes of climate forcing, if any, could account for occasional collapse of WAIS, I ran a number of million-year-long simulations. Million-year runs were necessary to dissipate the influence of an arbitrary initial condition and to allow an ensemble of ten 100,000-year glacial cycles to be generated by each experiment for statistical analysis. Longer runs were prohibitively expensive. I prescribed cyclic surface-temperature forcing, cyclic global sea-level forcing, or both. The initial condition in all experiments was the same: a model-generated steady-state ice-sheet configuration associated with glacial atmospheric and sea-level conditions (reference sea-level air temperature 10°C colder, and global sea level 120 m lower, than today).

I found that many combinations of surface-temperature and sea-level forcing could produce irregular ice-sheet fluctuations and sporadic collapse. This finding is best illustrated in Fig. 3 which shows the results of a simulation in which both the reference sea-level air temperature and global sea level were specified as the cyclic repetition of the simple sawtoothed 100,000-year time series shown in Fig. 1. Despite the simplicity and regularity of the climate forcing, the response of the ice-sheet is irregular. Three collapses occur (identified in Fig. 3 by the sea-level drawdown, a measure of ice-sheet size, going to zero) at $\sim 190,000$ years, $\sim 330,000$ years and $\sim 750,000$ years after the start of the simulation.

Response to transient change

To understand why the simulated history of WAIS discussed above is so irregular, it is constructive to examine the transient

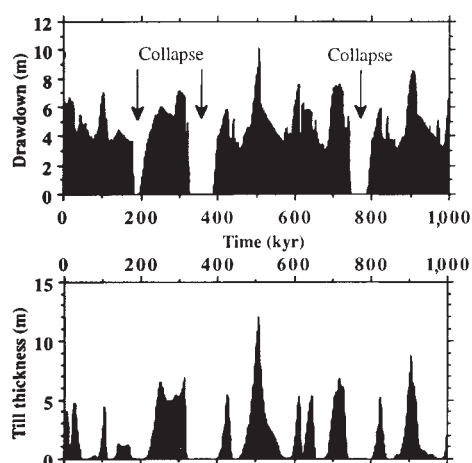


FIG. 3 Sea-level drawdown (upper panel) and average till thickness (lower panel) for a million-year simulation forced by the simple 100,000-year sawtoothed temperature and sea-level cycles in Fig. 1. Sea-level drawdown is a way of measuring the size of the WAIS. It represents the eustatic sea-level drop (in metres) required to build the particular configuration of the WAIS. This measure of ice-sheet size is zero when no grounded ice exists in West Antarctica and crustal rebound is complete. Ice-sheet collapse is thus signified when the sea-level drawdown is zero. The feedback effect of WAIS on sea level is not accounted for in the imposed external forcing of the ice sheet. Subglacial till thickness, averaged over the wet-bed fraction of the total ice-sheet footprint, is shown to verify model performance and to highlight the relation between till build-up and ice-sheet collapse. Typical till thickness observed below the West Antarctic ice streams today is 5–10 m (refs 9, 26).

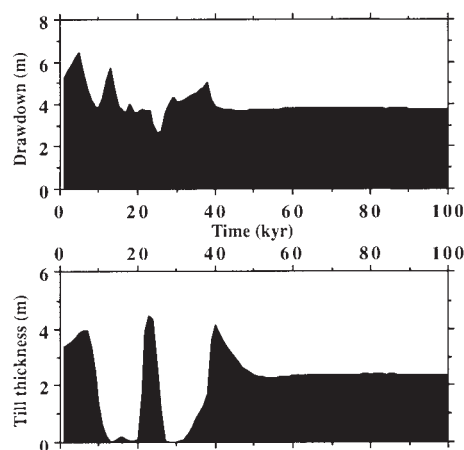


FIG. 4 Sea-level drawdown (upper panel) and average till thickness (lower panel) when the WAIS model is forced by a sudden switch between glacial and interglacial climate conditions.

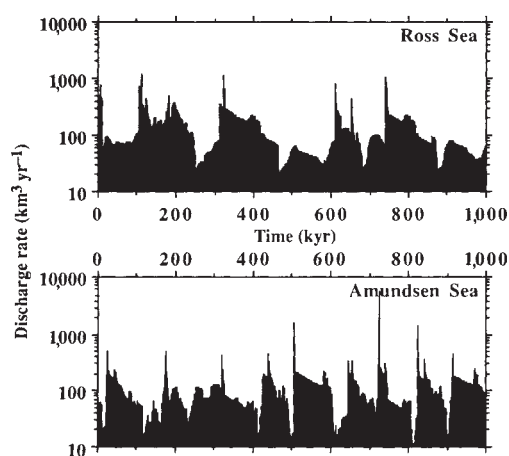


FIG. 5 Iceberg-discharge rates into the Ross (upper panel) and Amundsen Seas (lower panel) predicted by the million-year simulation of WAIS response to the simple sawtooth climate forcing shown in Fig. 1 (see also Fig. 3).

response to a sudden switch in climate from glacial to interglacial conditions. I will show that the timescale of this transient response is long enough to suggest that irregular behaviour described above is a simple consequence of mismatch between forcing timescale and the natural timescale of ice sheets which have deformable beds.

I simulated the ice sheet's evolution from an initial steady state (consistent with glacial conditions) used in the simulations discussed above towards a new steady state consistent with interglacial conditions (global sea level increased by 120 m and surface temperature increased by 10 °C over the glacial values). The WAIS was able to reach a stable steady state in this simulation, but only after a long period of adjustment. It took 50,000 years for the WAIS to equilibrate, and during this period there were several distinct oscillations of growth followed by rapid shrinkage (Fig. 4).

As seen by comparing the time series of average till thickness and the sea-level drawdown in Fig. 4, the slow build-up of deforming subglacial till towards a critical value seems to trigger rapid ice-sheet shrinkage. This is because ice-stream activity is generally dormant until the region containing deformable till connects with an open boundary through a frozen-bed zone that rings the margin of the ice sheet. Once this connection is established, ice-stream drainage into the Ross and Amundsen seas rapidly shrinks the ice sheet below what would eventually become its equilibrium size. Long timescales are introduced into this process because slow thermodynamic processes are involved in bringing frozen-bed zones up to the melting temperature.

The long timescale of WAIS adjustment to impulsive climate change (50 kyr) is comparable to the timescales of late-Pleistocene climate forcing (20, 40 and 100 kyr). In this circumstance, trends in WAIS configuration at any one instant in the past may not be in balance with the instantaneous state of external climate forcing.

Predictions

The simulations of WAIS evolution in response to a cyclic climate forcing suggest that the late Pleistocene history of the WAIS should be irregular. Verification of this irregularity may be possible by examining the geologic record for signs of sporadic collapse and periods of extreme iceberg discharge. Marine microfossil observations are consistent with the notion that collapse is a rare phenomena, and is not a periodic feature of the glacial cycle (ref. 6, and L. H. Buckle, personal communication). Figure 5 displays the iceberg discharge rates into the Ross and Amundsen seas associated with the million-year simulation discussed above (Fig. 3). Short periods of extreme iceberg discharge with amplitudes in excess of 500 km³ yr⁻¹ occur irregularly throughout the simulation. The occurrence of such extreme events may be detectable as prominent peaks in the ice-rafted debris abundance in offshore sediments, and perhaps as oxygen-isotope enrichments in local marine microfossils. I propose that the iceberg-discharge spikes simulated here may be West Antarctic equivalents of the Heinrich events thought to portray sudden, episodic iceberg discharge into the North Atlantic²⁴.

Another qualitative prediction gained from the simulations is the maximum rate of sea-level rise in response to WAIS collapse. The maximum rate for the simulation shown in Fig. 3 was ~0.25 m per century. Previous work using more realistic sub-ice topography and stronger climate forcing, but less-realistic ice-sheet physics, has suggested that larger rates of sea-level rise could be possible²⁵.

Conclusions

Model experiments indicate that climatic forcing associated with the glacial cycle could have elicited sporadic collapse and re-growth in the history of the WAIS. The irregularity of these events is a principal qualitative result which should be tested observationally by inspecting the marine record for evidence of sudden, strong iceberg discharges. On the prospect of forecasting WAIS response to CO₂-induced warming, I offer three opinions based on these experiments. First, collapse is not necessarily correlated with extremely warm or extremely cold periods of climate forcing. It is triggered by deformable till conditions at the bed of the ice sheet which, because it is isolated from the atmosphere, are affected by long-term climate trends (thousands of years) as well as the vagaries of ice-sheet history. Second, very long timescales (~50,000 years) can be involved in WAIS adjustment to simple, impulsive climate change because of the nonlinear nature of ice-stream dynamics. This suggests that forecasts of future behaviour will be sensitive to errors in the specification of (present) initial conditions which determine the ice sheet's predisposition towards rapid change. Third, the hypothesis that WAIS snow-accumulation rates increase when the climate warms does not contradict the hypothesis that future climate warming could cause the WAIS to collapse. My simulations incorporate the feedback between accumulation rate and climate warming, yet they display ice-sheet collapse or shrinkage at times when model climate forcing is warm. What is most important in determining the possibility for WAIS collapse in the near future is the distribution of deformable subglacial till.

Received 29 April; accepted 7 July 1992.

- Chappell, J. & Shackleton, N. J. *Nature* **324**, 137–140 (1986).
- Letreguilly, A. et al. *J. Glaciol.* **37**, 149–157 (1991).
- Hughes, T. *Rev. Geophys. Space Phys.* **13**, 502–526 (1975).
- Lambeck, K. & Nakada, M. *Nature* **357**, 125–128 (1992).
- Denton, G. H. et al. *Quat. Res.* **31**, 151–182 (1989).
- Scherer, R. P. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **90**, 395–412 (1991).
- Lorius, C., Raisbeck, G., Jouzel, J. & Raynaud, D. *The Environmental Record in Glaciers and Ice Sheets* (eds Oeschger, H. & Langway, C. C. Jr) 343–361 (Wiley, Chichester, 1989).
- Shackleton, N. J. *Quat. Sci. Rev.* **6**, 183–190 (1987).
- Alley, R. B. et al. *Nature* **322**, 57–59 (1986).
- Huybrechts, P. *Clim. Dynam.* **5**, 79–92 (1990).
- MacAyeal, D. R. *J. geophys. Res.* **94**, 4071–4087 (1989).
- MacAyeal, D. R. & Lange, M. A. *J. Glaciol.* **34**, 128–135 (1988).

- Huybrechts, P. *Ber. Polarforsch.* **99** (1992).
- Lindstrom, D. *Paleoceanography* **5**, 207–227 (1990).
- Lingle, C. S. & Clark, J. A. *J. geophys. Res.* **90**, 1100–1114 (1985).
- MacAyeal, D. R. & Thomas, R. H. *J. Glaciol.* **32**, 72–86 (1986).
- MacAyeal, D. R. *J. geophys. Res.* **97**, 595–603 (1992).
- Engelhardt, H. et al. *Science* **248**, 57–59 (1990).
- Kamb, B. *J. geophys. Res.* **96**, 16585–16595 (1991).
- Alley, R. B. et al. *J. geophys. Res.* **92**, 8931–8940 (1987).
- Boulton, G. S. & Hindmarsh, R. C. A. *J. geophys. Res.* **92**, 9059–9082 (1987).
- Alley, R. B. et al. *J. geophys. Res.* **92**, 8921–8929 (1987).
- Fortuin, J. P. F. & Oerlemans, J. *Ann. Glaciol.* **14**, 78–84 (1990).
- Broecker, W. et al. *Clim. Dynam.* **6**, 265–273 (1992).
- Thomas, R. H. et al. *Nature* **277**, 355–358 (1979).
- Blankenship, D. D. et al. *Nature* **322**, 54–57 (1986).

Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site

Gordon V. Louie*, Paul D. Brownlie*, Richard Lambert*, Jonathan B. Cooper*, Tom L. Blundell*, Steve P. Wood*, Martin J. Warren†, Sarah C. Woodcock† & Peter M. Jordan†

* Laboratory of Molecular Biology, Department of Crystallography, and Imperial Cancer Research Fund Structural Molecular Biology Unit, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

† School of Biological Sciences, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, UK

The three-domain structure of porphobilinogen deaminase, a key enzyme in the biosynthetic pathway of tetrapyrroles, has been defined by X-ray analysis at 1.9 Å resolution. Two of the domains structurally resemble the transferrins and periplasmic binding proteins. The dipyrromethane cofactor is covalently linked to domain 3 but is bound by extensive salt-bridges and hydrogen-bonds within the cleft between domains 1 and 2, at a position corresponding to the binding sites for small-molecule ligands in the analogous proteins. The X-ray structure and results from site-directed mutagenesis provide evidence for a single catalytic site. Interdomain flexibility may aid elongation of the polypyrrole product in the active-site cleft of the enzyme.

THE enzyme porphobilinogen deaminase (PBGD; EC 4.3.1.8, also referred to as hydroxymethyl bilane synthase or uroporphyrinogen I synthase) catalyses the stepwise polymerization of four molecules of porphobilinogen into the linear tetrapyrrole 1-hydroxymethylbilane, preuroporphyrinogen (for a review, see ref. 1). Preuroporphyrinogen is a key intermediate in the biosynthesis of haems, chlorophylls and corrins. PBGD occurs ubiquitously, with over 45% amino-acid sequence identity among the proteins from *Escherichia coli*², *Bacillus subtilis*³, *Euglena gracilis*⁴, rat⁵, mouse⁶ and man⁷. In humans, deficiency in the activity of PBGD is associated with the dominant hereditary disease, acute intermittent porphyria⁸.

Covalently linked to the side-chain of a conserved cysteine (Cys 242 in the 313 amino-acid protein from *E. coli*²) is a novel dipyrromethane cofactor, which is assembled by the monomeric apoenzyme from two molecules of the substrate porphobilinogen^{9,10}. The cofactor serves as a resident primer to which the polypyrrole chain is covalently attached during the tetrapolymerization reaction, but is not itself turned over. The polypyrrole is sequentially elongated by deamination of the incoming porphobilinogen substrate; carbon-carbon bond for-

mation through nucleophilic attack on the resultant methylene pyrrolinene by the free α -position of the terminal, enzyme-bound ring; and finally deprotonation at this position (Fig. 1). (For the first reaction in cofactor assembly, the nucleophile is a deprotonated cysteinyl sulphhydryl group.) The enzyme-intermediate complexes ES₁, ES₂, ES₃ and ES₄, generated during the course of the tetrapolymerization reaction, are stable to various degrees and can be isolated¹¹. Hydrolytic release of the product preuroporphyrinogen from ES₄ regenerates the holoenzyme, and occurs through reversal of the initial ring-coupling reaction (protonation at the α -position of the cofactor ring C2, cleavage of the carbon-carbon bond, and finally hydration of the tetrapyrrolic methylene pyrrolinene). PBGD catalyses all reaction steps stereospecifically^{12,13}.

Biochemical evidence suggests that all of the porphobilinogen-coupling steps occur at a single catalytic site¹¹. Thus, the mechanism by which the enzyme manoeuvres the growing polypyrrole chain to allow repeated use of the same catalytic machinery, and the structural factors that cause chain growth to be terminated at the tetrapyrrole stage are intriguing problems. We report here the three-dimensional structure of

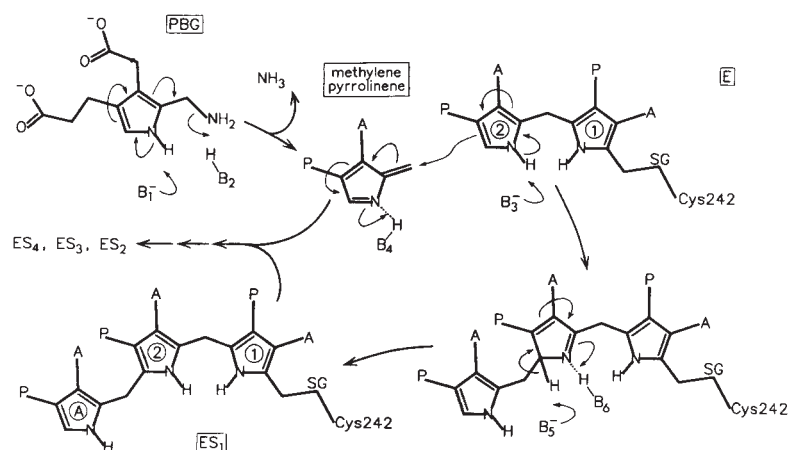


FIG. 1 Schematic representation of the dipyrromethane cofactor and of the generalized chemistry of a porphobilinogen ring-coupling reaction. Carbon-carbon bond formation proceeds through deamination (upper left) of the substrate porphobilinogen (PBG) to yield the methylene pyrrolinene; nucleophilic attack (upper right) by the carbon atom at the free α -position of the terminal, enzyme-bound ring; and deprotonation (lower right) at this same carbon atom. In the reaction depicted, the first ring (A) of the eventual product is added to ring C2 of the cofactor, thus converting the holoenzyme E into the enzyme-intermediate complex ES₁; generation of ES₂, ES₃, and ES₄ are analogous. The acetate and propionate side groups of the porphobilinogen moieties are denoted by A and P, respectively.

PBGD from *E. coli* elucidated using X-ray crystallographic methods to a resolution of 1.9 Å. The structure comprises three domains of the α/β class, two of which are structurally very similar to the transferrins and periplasmic binding proteins. We describe the interactions made by the novel dipyrromethane cofactor and discuss structural evidence for a single catalytic site and flexibly linked domains which may aid the repositioning of the dipyrromethane cofactor and enzyme-intermediate complexes during chain elongation.

Structural description

The structure of *E. coli* PBGD was determined by multiple isomorphous replacement (see legend to Fig. 2). The PBGD molecule has dimensions of about $57 \times 43 \times 32$ Å and the polypeptide chain is folded into three domains of roughly equal size (Fig. 3).

Domain 1 (residues 4–99 and 200–217) and domain 2 (105–193) have a similar overall topology, a modified, doubly wound parallel β -sheet¹⁴. Each sheet has four parallel and one antiparallel strands (domain 1: $-1x, +2x, +2x, -1$; domain 2: $-2, -1x, +2x, +2x$), with α -helical segments packed against each face and oriented roughly parallel to the strands. These two domains result from a duplication of a motif comprising a sheet of four parallel β -strands with intervening α -helices, followed by an additional separate strand. The duplicated motifs are then related by an approximate 2-fold axis (163°); the isolated strand of each motif interdigitates into the parallel β -sheet of the dyad-related motif, yielding the mixed sheet in each domain. Superposition¹⁵ of 53 pairs of equivalent α -carbons in the two domains yields a r.m.s. positional discrepancy of 1.5 Å. Both the strands and helices are shorter in domain 2, and there is little sequence identity except around the first α -helix in each domain (Ser 13=Ser 129, Gln 19=Gln 135, Pro 32=Pro 141). Domain 3 (residues 222 to the carboxy terminus) is an open-faced, three-stranded antiparallel β -sheet (+1, +1) with one face covered by three α -helices. The sheets in domains 1 and 2 are highly twisted (about 110° and 100° , respectively), whereas

the sheet in domain 3 is relatively flat (25° twist). Each domain contains a hydrophobic core, formed almost exclusively from the interdigitation of aliphatic side-chains projecting from the faces of the β -sheets and from the adjacent α -helices. (PBGD contains very few aromatic residues.) In domain 3, the hydrophobic core includes the interface between the two longer, roughly parallel helices and is formed mainly by a number of residues with small aliphatic side-chains (Ala 230, Ala 233, Ala 283, Gly 287, Ala 291).

Both the interdomain interactions and the connecting polypeptide-chain segments suggest that the domains are linked flexibly. There are few direct interactions between domains 1 and 2. The connecting segments between these two domains are formed by residues 100–104 and 194–199, and are roughly antiparallel. They interact through a number of hydrogen-bonds (102 N–199 O, Gln 198 N ϵ 2–104 O, Arg 101 N η 1–198 O), and each segment directly connects the C-terminal-edge strand of the origin domain to the antiparallel strand in the opposing domain ($\beta_{41/2}$ to $\beta_{52/1}$). It is possible that the two domains can hinge open about these connecting strands.

Domain 3 has small, hydrophobic interfaces with both domains 1 (Met 82, Ile 98, Gly 197/Ala 225, Val 228) and 2 (Leu 130, Cys 134, Leu 193/Pro 245, Leu 262, Pro 266, Gly 268). But the packing of domain 3 alongside the rest of the molecule is mediated primarily through polar interactions, which include: with domain 1, Arg 232 N η 1–85 O and Arg 232 N η 2–86 O; with domain 2, salt-bridges Arg 260–Glu 190 and Arg 273–Glu 138; and with the interdomain hinge region, 248 N–192 O, 101 N–Thr 224 O γ 1, Glu 197 N–Glu 231 O ϵ 2, and the salt-bridge Arg 101–Glu 231. Notably, site-directed substitutions at either of the invariant Arg 101 or Arg 232 inhibit enzyme activity^{16,17}. The interdomain region also contains several trapped water molecules. A single, short segment of polypeptide chain (residues 218–221) connects domains 1 and 3. Hinge-bending about this connecting segment would expose largely hydrophilic surfaces that could exist stably in an aqueous environment.

There is also evidence for flexibility of some loop regions.

TABLE 1 Summary of phasing analysis using heavy-atom derivatives of PBGD

Derivative	Soaking conditions	Merging <i>R</i> factor	Isomorphous <i>R</i> factor	Resolution (Å) and differences used	Cullis <i>R</i> factor	Overall phasing power	Number of sites	Binding site (and relative occupancy)
Chloroplatinite	10 mM 24 h	0.097	0.243	3.0 ID/AD	0.89	1.06	1	Pt (0.91): internal cavity near Arg 176
Uranyl fluoride	10 mM 24 h	0.166	0.294	3.0 ID/AD	0.69	1.83	4	U1 (0.97): Glu 163, Glu 256; U2 (0.55): Glu 72, Glu 292; U3 (0.61): Asp 254; U4 (0.31): Glu 88, Glu 239
Uranyl acetate	10 mM 24 h	0.122	0.345	3.0 ID/AD	0.71	1.86	3	U1 (0.84); U2 (0.83); U3 (0.97)
Uranyl sulphate	10 mM 20 h	0.045	0.197	3.3 ID	0.87	1.17	1	U3 (0.63)
<i>p</i> -Chloromercuri benzene sulphonate	10 mM 1 h	0.084	0.127	3.3 ID/AD	0.87	0.81	2	Hg1 (0.22): Cys 99, His 221, Lys 26; Hg2 (0.26): Cys 205, Arg 213, Val 96
Ytterbium chloride	10 mM 24 h	0.085	0.147	3.3 ID	0.88	0.84	1	U2 (0.47)

Merging $R = \sum_{hkl} \sum_{j=1}^n |I(hkl)_j - \bar{I}(hkl)| / \sum_{hkl} \sum_{j=1}^n I(hkl)_j$.

Isomorphous $R = \sum_{hkl} |F_{deriv}(hkl) - F_{nat}(hkl)| / \sum_{hkl} [F_{deriv}(hkl) + F_{nat}(hkl)]$.

ID, Isomorphous differences; AD, anomalous differences.

Cullis $R = \sum_{hkl} |F_{h0}(hkl) - F_{h0c}(hkl)| / \sum_{hkl} F_{h0}(hkl)$, considering only the centric reflections.

Overall phasing power = $[\sum_{hkl} F_{h0}^2(hkl) / \sum_{hkl} e^2(hkl)]^{1/2}$.

h, *k* and *l* are Miller indices defining crystal planes.

No well defined density is observed for residues 47–58, which form part of a loop connecting strand β_2 to helix α_2 and which have a solvent-accessible position at the front surface of the molecule. The high accessibility and mobility of this segment of polypeptide chain is consistent with the occurrence both of nearby sites of protease sensitivity and of the two most readily pyridoxylated lysines in the protein at residues 55 and 59^{11,18}.

Dipyrromethane cofactor and the active-site cleft

The dipyrromethane cofactor lies in a deep cleft between domains 1 and 2 (Fig. 3a). The floor and ceiling of the cleft are formed mainly from the C termini of the β -strands, the amino termini of α -helices, and the connecting loops. One face of the α_3 -helix and the hinge segments linking the two domains constitute the rear wall of the cleft. The cleft spans about half of both the width and depth of the molecule, and the positioning of the cofactor at its mouth leaves vacant a considerable volume of internal space, in which there are many well ordered water molecules.

Cys 242, which covalently binds ring C1 of the cofactor, is the second residue of a type I β -turn situated at the end of the loop joining α_1 and β_1 in domain 3. This loop projects into the cleft between domains 1 and 2, but does not form extensive interactions with either. The thioether linkage adopts a reasonably favourable 'left-handed spiral'¹⁴ conformation, very similar

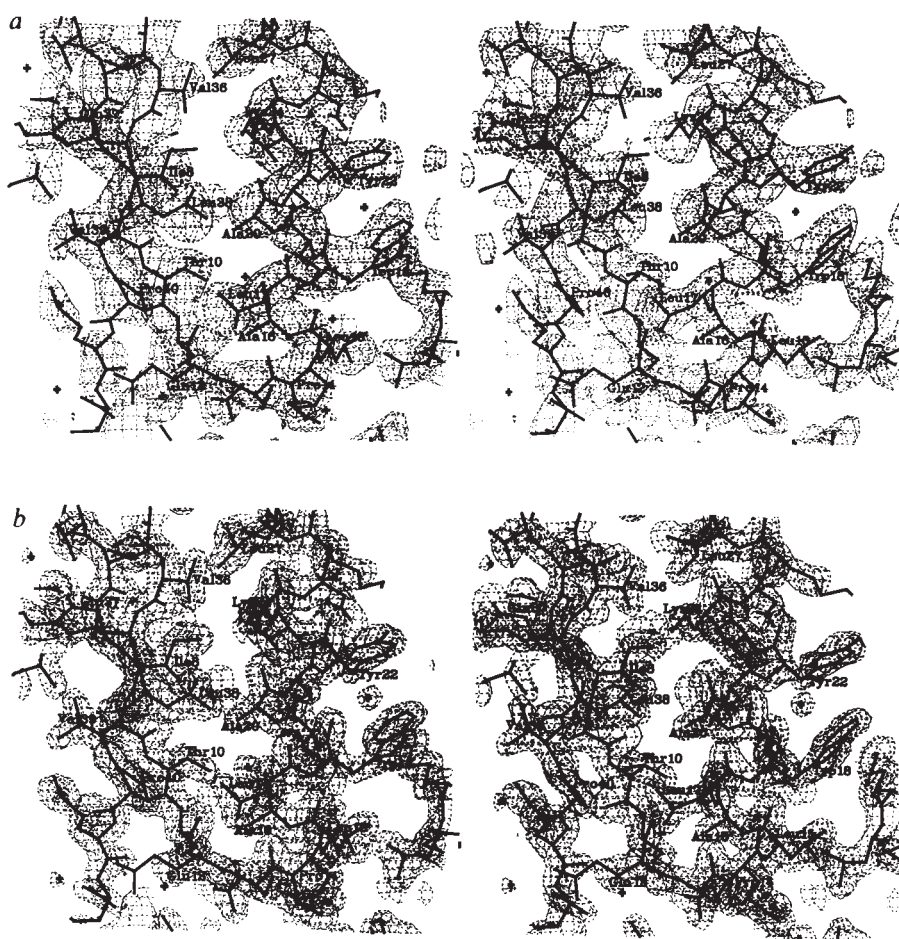
to that of the thioether to Cys 14 in cytochrome *c* (ref. 19); the two dihedral angles about the S γ atom have values of -80° and -70° , and χ_1 of the Cys 242 side chain is -60° . In all four acidic side groups of the dipyrromethane cofactor, the bond between the first and second atoms lies nearly perpendicular to the plane of the parent pyrrole ring (Fig. 4a). Additionally, the two propionate side groups have fully extended conformations.

The acetate and propionate side groups of the dipyrromethane form extensive salt-bridge and hydrogen-bond interactions with the surrounding protein matrix (mainly arginyl guanidiniums, seryl and threonyl hydroxyls, and backbone amide nitrogens) and bound water molecules (Fig. 4b). The amino-acid side-chain groups involved, which are notably invariant, participate in further hydrogen-bonding with adjacent protein and solvent atoms. Three of the carboxylate groups of the cofactor are positioned very near the N termini of α -helices. The two pyrrole nitrogens are both hydrogen-bonded to the side-chain carboxyl group of Asp 84. Furthermore, the pyrrole ring C2 is stacked against the phenyl ring of Phe 62. The cofactor's involvement in forming an extensive network of interactions, which cross-link the three domains, and its involvement in partially neutralizing the considerable electropositivity in the active-site cleft may explain the more stable and compact tertiary structure of the holoenzyme as compared to the apoenzyme^{11,20}.

A volume of electron density (Fig. 4a) situated ~ 4 Å behind

FIG. 2 Stereo diagrams of representative regions of the electron-density maps used in the structural determination of PBGD. *a*, The initial solvent-flattened MIR map at 3.0 Å resolution. *b*, The current $2F_o - F_c$ difference map. The region of the protein shown includes β_1 , α_1 and β_2 . The atomic skeleton is from the current refinement. Contouring for both maps is at 1.5 times the r.m.s. electron-density level.

METHODS. *E. coli* PBGD was prepared and crystallized from polyethylene glycol in acetate buffer (pH 5.0) as described previously²¹. The crystals belong to space group $P2_12_12$ with unit cell dimensions $a=88.0$, $b=75.9$ and $c=50.5$ Å. X-ray intensities for the native protein to 3.0 Å resolution and for five of the heavy-atom derivatives were measured on an Enraf-Nonius FAST area detector. The data set for the uranyl sulphate derivative was collected on a Siemens Xentronics area detector and for the native enzyme to 1.9 Å resolution on film at the Synchrotron Radiation Source at Daresbury. For the native data set, 180,969 independent measurements were reduced to 33,726 unique reflections with a merging R factor of 0.107 (on intensities), and completeness is greater than 99% over the entire resolution range ∞ –1.9 Å. Heavy-atom binding sites were located from inspection of difference Patterson and difference Fourier maps, and refined using the program PHARE. The phasing analysis (detailed in Table 1) gave a mean figure of merit of 0.73 for 6,315 reflections in the resolution range ∞ –3.0 Å. Solvent flattening³¹ significantly aided map interpretation, which was done using the program FRODO³² running on an Evans and Sutherland PS390 or on a Silicon Graphics IRIS-4D. Several iterations of refinement by restrained least-squares (RESTRAIN³³) or simulated annealing (X-PLOR³⁴), rebuilding against electron-density maps (coefficients $2F_o - F_c$ and $F_o - F_c$), and introduction of X-ray data of increasingly higher resolution have resulted in the current atomic model, in which the positions of residues 4–46 and 59–307, the dipyrromethane cofactor, 133 water molecules, and one acetate ion are determined. The crystallographic R factor for all 27,115 X-ray data in the resolution range 10.0–1.9 Å is 0.200 (0.184 for the subset of 24,379 reflections with $F \geq 3\sigma_F$ in the range 6.0–1.9 Å). The r.m.s. deviations from stereochemical ideality are 0.018 Å for bond distances, 0.032 Å for angle distances, 3.4° for ω peptide bonds, and 17° for side-chain dihedral angles.



No amino-acid residues have disallowed backbone conformations, according to the Ramachandran analysis³⁵. With the exception of the structural refinement, the CCP4 suite of programs (SERC Daresbury) has been used for all crystallographic calculations. A more detailed description of the structure determination will be given elsewhere, after further refinement. The current atomic coordinates will be deposited with the Brookhaven Protein Data Bank³⁶.

the observed position of the cofactor and emanating from a weak bifurcation in the density of the bridging methylene group almost certainly represents an alternative position for the second (C2) pyrrole moiety. (A free acetate ion (presumably from the solvent) and a water molecule have been modelled into the

density corresponding to the positions of the carboxylate groups of the C2 ring in its alternative conformation.) However, the dipyrromethane cofactor existed largely in an oxidized form in the crystal samples used for the X-ray analysis²¹. It is thus possible that the position of ring C2 observed with the oxidized

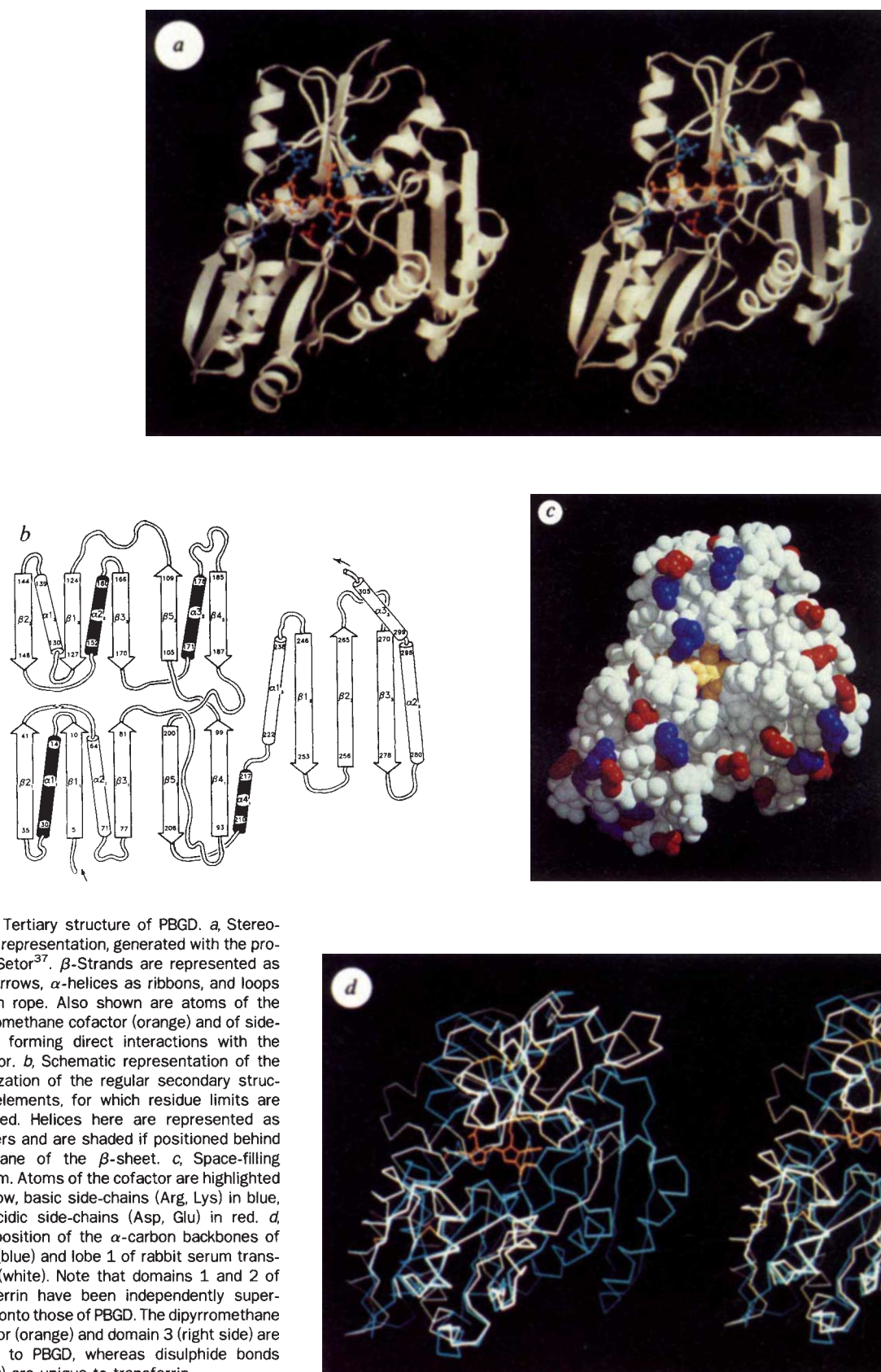


FIG. 3 Tertiary structure of PBGD. *a*, Stereo-ribbon representation, generated with the program Setor³⁷. β -Strands are represented as wide arrows, α -helices as ribbons, and loops as thin rope. Also shown are atoms of the dipyrromethane cofactor (orange) and of side-chains forming direct interactions with the cofactor. *b*, Schematic representation of the organization of the regular secondary structural elements, for which residue limits are indicated. Helices here are represented as cylinders and are shaded if positioned behind the plane of the β -sheet. *c*, Space-filling diagram. Atoms of the cofactor are highlighted in yellow, basic side-chains (Arg, Lys) in blue, and acidic side-chains (Asp, Glu) in red. *d*, Superposition of the α -carbon backbones of PBGD (blue) and lobe 1 of rabbit serum transferrin (white). Note that domains 1 and 2 of transferrin have been independently superposed onto those of PBGD. The dipyrromethane cofactor (orange) and domain 3 (right side) are unique to PBGD, whereas disulphide bonds (yellow) are unique to transferrin.

form of the cofactor actually corresponds very nearly to the position at which the substrate porphobilinogen binds in the catalytic site. Moreover, the more internal, alternative position may represent the genuine binding site for ring C2. This proposed disposition of ring C2 with respect to the catalytic site

rationalizes the stability of the dipyrromethane cofactor toward catalytic turnover^{9,11}. Of the three identified binding sites for pyrrole moieties, the greatest number of specific interactions are formed at that occupied by ring C1 of the cofactor, consistent with the relatively low atomic temperature factors of this group.

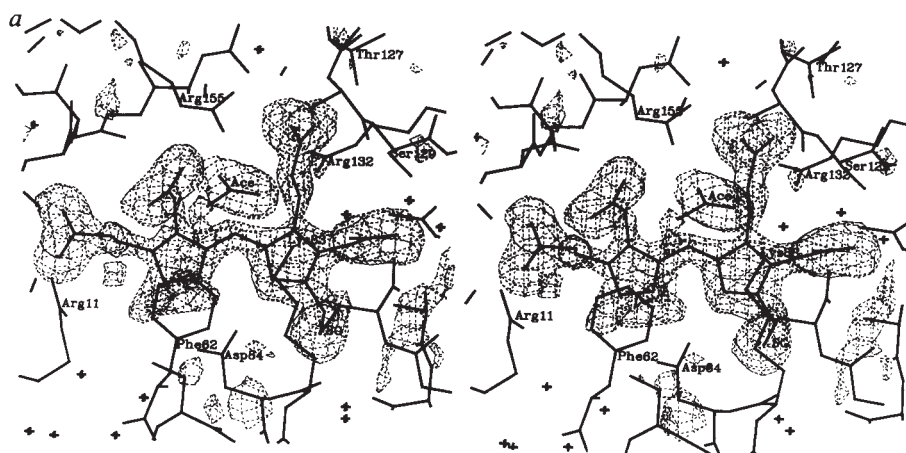
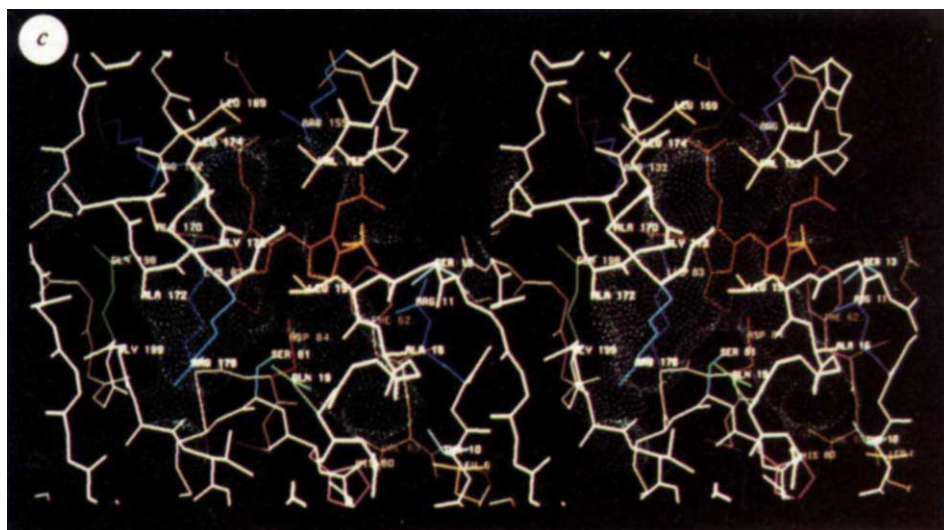
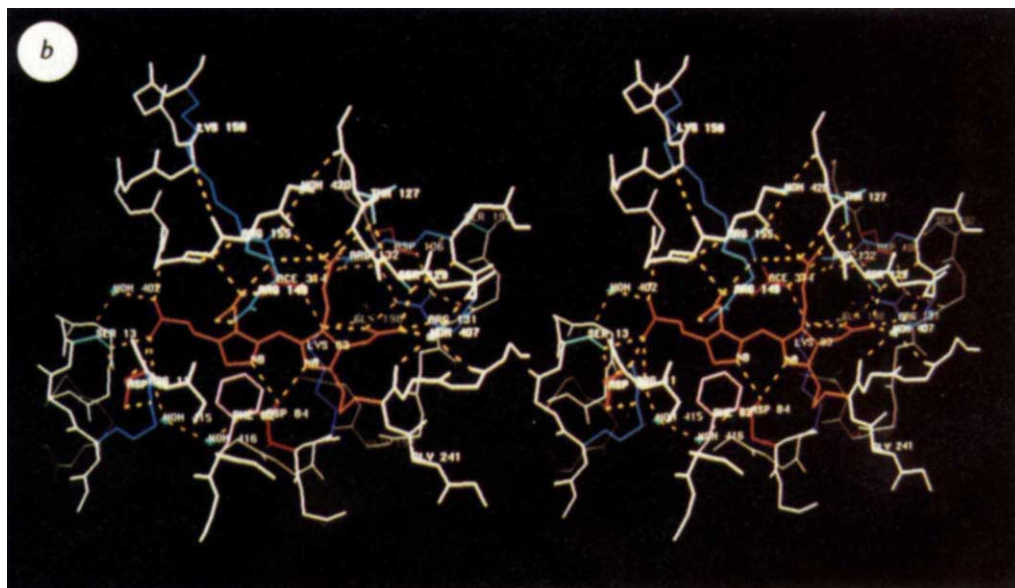


FIG. 4 The dipyrromethane cofactor. *a*, Stereo view of electron density for the cofactor from a $F_o - F_c$ omit map. Contouring is at 2.0 times the r.m.s. level. An excess lobe of electron density connected to the free α -position (lower left) of ring C2 is consistent with the presence of a carbonyl-oxygen atom and suggests the cofactor existed, at least partially, in an oxidized (dipyrromethanone or dipyrromethenone) form in the crystals studied. But because both porphobilinogen units adopt apparently favourable conformations and form numerous stabilizing interactions with the protein molecule (see text), the position occupied by the cofactor undoubtedly represents a functionally important binding site. *b*, Protein-cofactor interactions. Hydrogen-bonds and salt-bridges are drawn as broken yellow lines. *c*, Dot surface³⁸ of the internal cavity located behind the cofactor. Amino-acid residues lining the cavity are labelled.



Structural similarity with the binding proteins

An intriguing structural relationship between PBGD and several binding proteins further supports the idea that hinge bending between domains 1 and 2 is important in aiding ligand binding in the interdomain cleft. The overall topology of domains 1 and 2 of PBGD is identical to that of the group-II periplasmic binding proteins²² (including sulphate⁻²³, phosphate⁻²⁴ and maltodextrin⁻²² binding proteins) and of the duplicated lobe of the bilobal transferrins^{25,26}: among all of these proteins, the connectivity in the two α/β -domains is the same and there are two interdomain hinge segments (Fig. 3b). Also similar are the overall twists of the core β -sheets and the lengths of corresponding secondary-structural elements. There is, however, significant variability in the lengths of loops connecting the alternating β -strand and α -helical segments (these are generally shortest in PBGD), in the precise lateral positioning of the α -helices with respect to the faces of the β -sheets, and in the relative orientation of domains 1 and 2 in each protein molecule. Superposition, considering domains 1 and 2 independently, of the α -carbon backbones of PBGD and lobe 1 of rabbit serum transferrin²⁵ (entry 1TFD in the Brookhaven Protein Data Bank) yields an overall r.m.s. positional deviation of 1.9 Å for 115 equivalent α -carbons (Fig. 3d).

A common feature of PBGD and the binding proteins is a binding cleft at the interface between the two domains. The N terminus of one (α_1) or more α -helices is involved in forming the binding site, and with the exception of maltodextrin-binding protein, the ligand is oxyanionic (Table 2). The binding cleft in PBGD has a considerably greater volume than those in the binding proteins. This not only reflects the larger size of the dipyrromethane cofactor, but also indicates that the cleft may have evolved to accommodate the tetrapyrrolic product. In comparison to transferrin, the larger cleft-volume in PBGD results from a shift in the relative positions of domains 1 and 2, which also orients the helices more optimally for interaction with the carboxylate groups of the cofactor (and also of the substrate). Although there is no extensive sequence homology between PBGD and the other proteins, several functionally important residues in roughly equivalent positions are conserved (Table 2). They include a serine that both caps the α_1 -helix and hydrogen-bonds an acidic oxygen on the ligand; an arginine near the N terminus of this helix which directly or indirectly interacts ionically with the oxyanion; and an aspartate on the loop joining β_3 and α_3 , which has diverse roles.

The transferrins undergo a hinge-bending motion during binding first of the bicarbonate anion and then of the ferric ion²⁷. This cleft opening aids access of the ligands to their binding sites, and subsequent closure creates an environment in which the bound ligands are excluded from solvent. A corresponding hinge-bending in PBGD implies opening of the active-site cleft, possibly accompanied by movement of domain 3 away from domains 1 and 2. This would allow the cofactor and developing product to reposition on binding of each incoming molecule of porphobilinogen substrate. Such a mechanism could also allow the intermediates to attain optimal binding interactions at each stage of the chain elongation.

Implications for the mechanism of PBGD

Two aspects of the polymerization of porphobilinogen by PBGD must be considered: the mechanism of catalysis of the bond formations, and the means by which the growing polypyrrole chain is manoeuvred in the active site of the enzyme. The determination of the three-dimensional structure of PBGD has ruled out the unlikely possibility that the individual ring-couplings are catalysed at (up to four) distinct sites, as there are no signs of replicated catalytic machinery. This agrees with biochemical evidence indicating the presence of only a single catalytic site¹¹.

As in the nonenzymic reaction, acid catalysis²⁸ is presumably involved in the PBGD-catalysed reaction (as illustrated in

Fig. 1). The side-chain of Asp 84 is close to both pyrrole nitrogens of the cofactor (Fig. 4). In its solvent-inaccessible position behind the phenyl ring of Phe 62, it would be expected to be protonated and could thus serve as a suitable acid group in promoting catalysis. In addition, once deprotonated, the Asp 84 carboxylate group is ideally positioned to stabilize the positive charges on the pyrrole nitrogens of the condensing rings. Our recent results from site-directed mutagenesis, in which Asp 84 has been replaced by Glu, Asn and Ala, confirm that this residue is essential for catalytic activity (S.C.W. *et al.*, manuscript in preparation). Asp 84 corresponds to a functionally important group in the analogous transferrins and periplasmic binding proteins, and is evolutionarily conserved.

The observed cofactor-protein interactions, as discussed above (Fig. 4b), combined with results from arginine mutagenesis^{16,17} have defined possible porphobilinogen-binding sites in the active site. Replacement of Arg 131 or Arg 132, which are involved in forming site C1, completely prevents assembly of the cofactor. Both Arg 149 and Arg 11 form part of site C2, and Arg 155 forms part of sites C1 and C2. Replacement of these residues (particularly the latter two) severely inhibits substrate binding and chain elongation. It can be proposed that after the cofactor has been assembled, each subsequent incoming porphobilinogen molecule binds at site C2, displacing the formerly resident pyrrole ring. This scheme is consistent with experiments on enzyme-intermediate complexes showing that in the absence of added porphobilinogen substrate, the terminal ring of the polypyrrole chain (presumably occupying the catalytic site) is susceptible to hydrolysis and release as hydroxy-porphobilinogen¹¹. Possibly the displaced ring assumes the more internal position for the cofactor ring C2 evident in electron density maps (see above), and thus remains sufficiently close to react with the incoming substrate. An additional point is that the absence of acidic side-chains immediately surrounding the entrance to the active-site cleft (Fig. 3c) may aid entry of the anionic substrate molecule.

Clearly if PBGD has only a single catalytic site, it must reposition the polypyrrole chain after each ring-coupling step so that the terminal pyrrole unit can react further. The X-ray structure and biochemical data suggest two possible models. In the first, the chain fills the large, internal cavity located behind the cofactor (Fig. 4c). This model provides an obvious, steric mechanism for preventing chain growth beyond a specific number (four) of porphobilinogen units. Such a model also explains the intriguing conservation of residues lining the cavity (in particular the small side-chains Ala 170, Ala 172 and Gly 173 which occur on the inward face of the α_3 -helix), and the observation that replacement of Arg 176, which is positioned at the rear of the cavity, inhibits the formation of ES₃ from ES₂^{16,17}. Alternatively, the polypyrrole chain may be pulled through the catalytic site (such that the binding sites C1 and C2 always hold the penultimate and terminal rings, respectively), possibly by rigid-body movement of all or part of domain 3. The reactivity of Cys 134 (which occurs near the interface between domains 2 and 3) with *N*-ethylmaleimide increases progressively as PBGD proceeds from the holoenzyme E through to ES₃¹¹. This result suggests that the positioning of domain 3 changes over the course of the complete reaction cycle, and may be most consistent with the second model, although the first also probably requires some domain movement to accommodate the growing product. Further site-directed mutagenesis studies and structural analyses of both enzyme-intermediate and enzyme-inhibitor complexes will undoubtedly clarify the precise mechanism by which the growing polypyrrole chain is manipulated by the enzyme.

Whatever the precise course taken by the polypyrrole chain as it is extended in the active site of the enzyme, considerable local flexibility is probably necessary to permit entry of the substrate, conformational adjustments associated both with attachment of the cofactor and with extension of the pyrrole

TABLE 2 Some characteristics of the binding clefts in PBGD and the binding proteins

Protein	PBGD	SBP	TFN	PBP	MBP
Ligand	Carboxylate side groups on porphobilinogen	Sulphate	Bicarbonate	Phosphate	Maltodextrin, maltose
Binding helices	$\alpha 1_1, \alpha 1_2, \alpha 2_2$	$\alpha 2_1, \alpha 1_2, \alpha 2_2$	$\alpha 1_2$	$\alpha 2_1, \alpha 1_2, \alpha 2_2$	—
$\alpha 1_2$ -Serine	Ser 129	Ser 130	Thr 120?	Ser 139	—
$\alpha 1_2$ -Arginine	Arg 131	Arg 134	Arg 124	Arg 135	—
$\beta 3_1/\alpha 3_1$ -Aspartate	Asp 84	Asp 68	Asp 63	Asp 56	Asp 65
Role of aspartate side-chain	Bind pyrrole nitrogens; catalysis?	Salt-bridge Arg 134	Coordinate iron ion	Hydrogen-bond phosphate	Hydrogen-bond sugar hydroxyl

chain, and finally release of the preuroporphyrinogen product. Interdomain hinge-bending as discussed above may provide one source of such flexibility. In addition, flexibility is apparent in the loop of polypeptide chain carrying the cofactor: this loop makes few interactions with the rest of the molecule and has relatively high temperature factors, and Cys 242 is preceded by two glycine residues. Flexibility in the extended acetate and propionate side groups may also be important in maintaining favourable interactions between the cofactor (or the polypyrrole chain) and the protein throughout the course of product elongation. Finally, the stretch of polypeptide chain (containing residues 47–58) passing directly in front of the active site cleft appears to be highly mobile. It may constitute a flexible 'lid' with a role in transiently blocking access of solvent to the active site during catalysis, which may be important both for protecting reactive intermediates and for modulating the nature of the catalytic environment. This proposed role may explain the lowered activity of site-directed mutants at Lys 55 and Lys 59 and also the observed, substrate-elicited protection of these residues from pyridoxylation^{11,18}.

Evolutionary aspects

Sequence comparisons indicate that the PBGDs for which primary structures are known (including human) share a common three-dimensional structure and mechanism of activity. Insertions in the sequence of the other species of PBGD relative to the *E. coli* protein occur at surface loops, which are structurally reasonable positions. The 29 extra residues in the higher eukaryotic sequences relative to the bacterial sequences occur in the loop either preceding or following strand $\beta 3_3$ in domain 3. Most of the invariant residues in the known PBGD sequences are clustered around the active-site cleft. These are involved in forming interactions with the cofactor and/or promoting growth of the polypyrrole chain (see above). Many of the invariant

residues more remote from the active-site cleft appear to be of structural importance, and have roles in stabilizing turn conformations, forming the hydrophobic core of the molecule, or forming capping interactions at the ends of helices. Glycine or alanine frequently occupy locations of restricted steric space, and at several positions in the polypeptide chain a conformation inaccessible to residues bearing a β -carbon is adopted by an invariant glycine.

A number of amino-acid replacements in human PBGD have been reported that give rise to acute intermittent porphyria^{29,30}. With the structure determination of *E. coli* PBGD, the possible structural and functional effects of these mutations can be interpreted. For example, the evolutionarily invariant Gln 19, Arg 149, and Arg 155 are involved in a network of hydrogen-bonding interactions within the active-site cleft (Fig. 4b, c), which would be disrupted in mutant proteins (with Lys 19, Gln 149, or Gln 155 substitutions). Most notably, the two arginine residues interact directly through salt-bridges with the acidic side groups of the cofactor, and may form part of the binding site for the porphobilinogen substrate. Leu 159, also invariant, is packed in the hydrophobic core of the protein molecule, and replacement by arginine would be structurally deleterious.

The close structural resemblance among PBGD and the binding proteins suggests these proteins are related by divergent evolution. It can be speculated that the PBGD enzyme originated, through recruitment of catalytic groups and perhaps the mobile lid (which is absent in all of the binding proteins), from a two-domain binding protein specific for an anionic, porphobilinogen-like substrate. Furthermore, the ancestral PBGD may have rudimentally catalysed the coupling of porphobilinogen molecules within its binding cleft without the aid of the covalently bound cofactor, which arose later when the C-terminal domain was acquired. □

- Jordan, P. M. in *New Comprehensive Biochemistry* Vol. 19 (eds Neuberger, A. & Van Deenen, L. L. M.) 1–66 (Elsevier, Amsterdam, 1991).
- Thomas, S. D. & Jordan, P. M. *Nucleic Acids Res.* **14**, 6215–6226 (1986).
- Petricek, M., Rutberg, L., Schroder, I. & Hederstedt, L. *J. Bact.* **172**, 2250–2258 (1990).
- Sharif, A. L., Smith, A. G. & Abell, C. *Eur. J. Biochem.* **184**, 353–359 (1989).
- Stubnicer, A. C., Picat, C. & Grandchamp, B. *Nucleic Acids Res.* **16**, 3102 (1988).
- Beaumont, C. *et al. J. Biol. Chem.* **264**, 14829–14834 (1989).
- Raich, N. *et al. Nucleic Acids Res.* **14**, 5955–5968 (1986).
- Kappas, A., Sassa, S., Galbraith, R. A. & Nordmann, Y. in *The Metabolic Basis of Inherited Disease* (eds Scriver, C. R., Beaudet, A., Sly, W. S. & Valle, D.) 1305–1365 (McGraw-Hill, New York, 1989).
- Jordan, P. M. & Warren, M. J. *FEBS Lett.* **225**, 87–92 (1987).
- Hart, G. J., Miller, A. D., Leeper, F. J. & Battersby, A. R. *J. chem. Soc. chem. Commun.* **23**, 1762–1765 (1987).
- Warren, M. J. & Jordan, P. M. *Biochemistry* **27**, 9020–9029 (1988).
- Jones, C., Jordan, P. M. & Akhtar, M. A. *J. chem. Soc. Perkin Trans.* **1**, 2625–2633 (1984).
- Schneider, J.-R., Jendrezewski, S., Abell, A., Hart, G. J. & Battersby, A. R. *J. chem. soc. Chem. Commun.* **6**, 436–438 (1987).
- Richardson, J. S. *Adv. Protein Chem.* **34**, 167–339 (1981).
- Sutcliffe, M. J., Haneef, I., Carney, D. & Blundell, T. L. *Protein Engng* **1**, 377–384 (1987).
- Lander, M. *et al. Biochem. J.* **275**, 447–452 (1991).
- Jordan, P. M. & Woodcock, S. C. *Biochem. J.* **280**, 445–449 (1991).
- Hadener, A., Alefounder, P. R., Hart, G. J. & Battersby, A. R. *Biochem. J.* **271**, 487–491 (1990).
- Louie, G. V. & Brayer, G. D. *J. molec. Biol.* **214**, 527–542 (1990).
- Scott, A. I. *et al. FEBS Lett.* **242**, 319–324 (1989).
- Jordan, P. M. *et al. J. molec. Biol.* **224**, 269–271 (1992).
- Spurlino, J., Lu, G.-Y. & Quirocho, F. A. *J. Biol. Chem.* **266**, 5202–5219 (1991).

- Pflugrath, J. W. & Quirocho, F. A. *J. molec. Biol.* **200**, 163–180 (1988).
- Luecke, H. & Quirocho, F. A. *Nature* **347**, 402–406 (1990).
- Baker, E. N., Rumball, S. V. & Anderson, B. F. *Trends biochem. Sci.* **12**, 350–353 (1987).
- Sarra, R., Garratt, R., Gorinsky, B., Jhoti, H. & Lindley, P. *Acta Crystallogr.* **B46**, 763–771 (1990).
- Anderson, B. F. *et al. Nature* **344**, 784–787 (1990).
- Mauzerall, D. *J. Am. chem. Soc.* **82**, 2605–2609 (1960).
- Delfau, M. H. *et al. J. clin. Invest.* **86**, 1511–1516 (1990).
- Mgone, C. S., Lanyon, W. G., Moore, M. R. & Connor, J. M. *Hum. Genet.* (in the press).
- Wang, B. C. *Meth. Enzym.* **115**, 90–111 (1985).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268–272 (1978).
- Haneef, I., Moss, D. S., Stanford, M. J. & Borkakoti, N. *Acta Crystallogr.* **A41**, 426–433 (1985).
- Brunger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
- Ramakrishnan, C. & Ramachandran, G. N. *Biophys. J.* **5**, 909–933 (1965).
- Bernstein, F. C. *et al. J. molec. Biol.* **112**, 533–542 (1977).
- Evans, S. V. *J. molec. Graphics* (in the press).
- Connolly, M. L. *Science* **221**, 709–713 (1983).

ACKNOWLEDGEMENTS. We thank I. Tickle and S. Evans for assistance with computer software, R. Pickersgill of the AFRC in Reading for measuring the data set on the uranyl sulphate derivative and SERC Daresbury for use of data collection facilities. J. Thornton initially suggested the possibility of structural resemblance between PBGD and transferrin. F. Leeper first noted that electron density for the cofactor indicated a dipyrromethenone form. R. Bonnett contributed useful discussions on the properties of dipyrromethenones. P. Lindley and R. Sarra on the comparison between PBGD and transferrin and W. Lanyon on mutations causing acute intermittent porphyria. This research was funded by the SERC (Molecular Recognition Initiative) and the AFRC (P.M.J.). G.V.L. is the recipient of an MRC Canada Postdoctoral Fellowship.

A photoionization instability in the early intergalactic medium

Craig J. Hogan

Astronomy and Physics Departments, University of Washington, Seattle, Washington 98195, USA

THE distribution of neutral hydrogen absorption lines in quasars indicates that the intergalactic medium is fully formed even at the highest redshifts^{1,2}: highly ionized gas fills most of space, but there are many denser, lower-entropy condensations with a higher fraction of neutral gas, and a few very dense, optically thick clouds thought to be associated with the progenitors of modern galaxies. This early ionization requires there to have been some form of energy injection (heat or light) at redshift $z > 5$, but its precise origin is unknown. (Possible sources of ionizing energy range from decaying massive neutrinos³⁻⁵ to faint but numerous active galaxies.) I argue here that any fairly uniform source of ionizing photons can be the cause of an instability in the pre-galactic medium on scales larger than a photon path length. Underdense regions receive more ionizing energy per atom and reach higher temperature and entropy, driving the density down still further. Fluctuations created by this instability can lead to the formation of structures resembling protogalaxies and intergalactic clouds, obviating the need for gas clouds or density perturbations of earlier cosmological provenance, as is usually assumed in theories of galaxy and structure formation.

In general the ionization problem has been thought to be separate from the problem of the origin of galaxies and intergalactic clouds. The photoionization instability described here, however, connects the two issues by allowing the spontaneous generation of inhomogeneity in the early cosmic medium. Unlike other conjectural post-recombination instabilities^{6,7}, this instability requires no 'special' opacities or photon spectra, but generically accompanies the process of photoionization. The essential physical elements of the model belong to a broad class of well-known coupled acoustic/ionization/thermal instabilities which occur in various forms in the interstellar medium of the Galaxy⁸⁻¹⁰. At the very low density of intergalactic gas, the familiar galactic modes grow too slowly to be important. But under the different conditions assumed to hold in the pre-galactic Universe—in particular, the precise uniformity of the gas and photon source, the low density and the long cooling time—new modes of instability assume prominence and naturally create the observed two-phase structure. The presence of a simple and generic instability of pre-galactic ionization naturally suggests that the reionization of the Universe may have caused the formation of gas clouds.

Let us consider a medium of pure hydrogen gas with atomic mass m_H , density $\rho = n_H m_H$, pressure p , temperature T and neutral fraction x . Embedded in this gas is a source of ionizing photons, not coupled hydrodynamically to the gas; it remains uniform even if the gas is perturbed (except for gravitational response, which is slow enough to be negligible). Let ω_i denote the characteristic rate for the medium to ionize completely, defined as the ratio of the photon production rate to the total atomic number density n_H . Let $\omega_{\text{rec}} = n_H(1-x)\alpha$ denote the recombination rate per ion, where α is the recombination coefficient. In static ionization equilibrium, $\omega_i = (1-x)\omega_{\text{rec}}$; thus if $\omega_i < \omega_{\text{rec}}$, the neutral fraction adjusts itself accordingly to achieve equilibrium, whereas if $\omega_i > \omega_{\text{rec}}$, the neutral fraction steadily decreases with time, asymptotically approaching zero. Typically in cosmic ionization models at early times $\omega_i < \omega_{\text{rec}}$, yielding a mostly neutral medium; the bulk of the gas becomes ionized when $\omega_i \approx \max[H, \omega_{\text{rec}}]$, where H is the expansion rate.

Now consider perturbations in the gas, leaving the ionizing source unperturbed. Let $\delta\rho$, δp , δT denote the fluctuation from the mean values $\bar{\rho}$, $\bar{p}$, $\bar{T}$, and $\delta \ln \rho$, $\delta \ln p$, $\delta \ln T$ their fractional

perturbations. The dependence of δp on $\delta\rho$ determines the behaviour of perturbations. In a static background with no gravity, the linearized hydrodynamic equations give the time behaviour of plane-wave fluctuations with spatial wavevector $\mathbf{k}$: $\delta \ln T \propto \delta \ln p \propto \delta \ln \rho \propto e^{-i(\omega t - \mathbf{k} \cdot \mathbf{x})}$, where

$$\omega^2 = (\partial p / \partial \rho) k^2 \quad (1)$$

For perturbations of adiabatic gas, $(\partial p / \partial \rho)$ is positive, leading to propagating acoustic oscillations with adiabatic sound speed $v_s^2 \equiv (\partial p / \partial \rho)_{\text{adiabatic}} = \gamma((2-x)kT/m_H)$, where $\gamma \equiv (\partial \ln p / \partial \ln \rho)_{\text{adiabatic}} = 5/3$ denotes the adiabatic index for an ideal gas.

Here, however, the pressure is also perturbed non-adiabatically because of the temperature perturbation caused by photon heating. We estimate $\delta \ln T$ assuming a cooling rate much slower than $|\omega|$. (This is reasonable during the early stages of ionization, until the medium is heated by ionization to $T \approx 10^4$ K, and if the redshift and ionization are low enough for Compton cooling to be inefficient.) Each photoelectron then simply increases the heat content of the gas by $h\nu - \text{Ryd}$, where ν is the frequency and Ryd is the photoionization threshold energy for hydrogen (13.6 eV). The perturbation in heating gives the time derivative of the temperature perturbation, $k\delta\dot{T} = -i\omega k\delta T = E_{\text{ev}}\delta\omega_i$, where $E_{\text{ev}} \equiv (h\nu - \text{Ryd})[2/3(2-x)]$ denotes the change in kT per particle caused by each photoejection, giving $\delta \ln T = i(E_{\text{ev}}/kT)(\omega_i/\omega)\delta \ln \omega_i$. For optically thick gas, all the photons are consumed by photoabsorption, so $\delta \ln \omega_i = -\delta \ln \rho$. (This follows for example from considering a uniform source of photons per volume per time, n_X/τ_X , with fixed n_X , giving $\tau_X\omega_i = n_X/n_H$.) Because $\delta \ln p = \delta \ln T + \delta \ln n$, and ignoring $\text{Im}(\delta \ln n)$

$$\frac{\delta \ln p}{\delta \ln \rho} \approx \gamma - i \frac{E_{\text{ev}} \omega_i}{kT \omega} \quad (2)$$

Equations (1) and (2) then lead to a cubic dispersion relation for the perturbations

$$\omega^3 - \omega(v_s k)^2 + i(E_{\text{ev}}/\gamma kT)\omega_i(v_s k)^2 = 0 \quad (3)$$

The unstable mode is the positive imaginary root of equation (3). Pressure and density perturbations both grow exponentially, with the amplitude ratio given by equation (2).

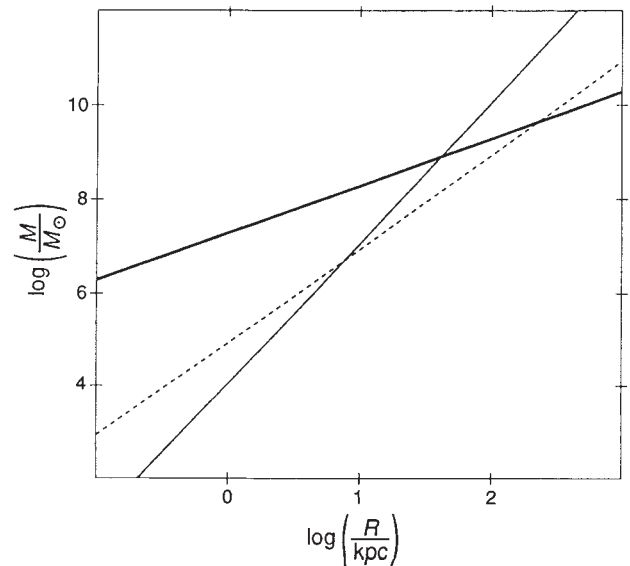


FIG. 1 Characteristic masses for clouds produced by the instability, with $\log$ mass in solar units plotted against $\log$ radius in kiloparsecs (kpc). The dashed line shows the minimum mass for the clouds M_{min} for $h\nu = 2$ Ryd, neutral fraction $x = 0.5$. The heavy line shows M_{jeans} , for $T = 10^4$ K. The light solid line corresponds to the mass within a sphere of radius R with constant density ρ , shown here for $n_H = 10^{-4} \text{ cm}^{-3}$. Systems form along this line, pressure-confined condensations (absorption clouds) below M_{jeans} , gravitationally collapsed systems (galaxies) above it.

For large scales (small k), the growth rate is faster than the sound frequency ($v_s k$), and

$$\omega \approx i[(E_{e\gamma}/\gamma kT)\omega_i(v_s k)^2]^{1/3} \quad (4)$$

On small scales, the growth rate is

$$\omega \approx i(E_{e\gamma}/\gamma kT)\omega_i \quad (5)$$

independent of scale; here the instability is almost isobaric ($\delta \ln p \approx 0$), and the perturbations in density and entropy grow close to pressure equilibrium.

An important effect not described in these equations is that on length scales small enough to be optically thin, the instability growth rate decreases, because a density perturbation no longer creates a corresponding heating perturbation: $|\delta \ln \omega_i| \ll |\delta \ln \rho|$. Because the growth rate decreases on small scales, the instability makes coherent astronomical structures, rather than merely causing microscopic turbulence and heating.

At the epoch of cosmic ionization $\omega_i \geq H$. In general (although not always^{3,4}) the photoejection energy $E_{e\gamma}$ is at least of the order of Rydbergs (tens of electron volts), much greater than the thermal energy $kT = 1 \text{ eV} (T/10^4 \text{ K})$. Thus, in spite of the approximations used here, it is clear that the process of cosmic photoionization is often accompanied by an instability with $-\dot{\omega} \gg H$ over some range of scales. The instability grows roughly exponentially at a rate much faster than the typical evolution rates of cosmological models, and linear theory predicts that the fluctuations become nonlinear. The statistical properties of the resulting nonlinear structures are insensitive to initial perturbations, but depend mainly on the gas dynamics.

The instability turns off once fluctuations become nonlinear, because the source of the ionizing radiation becomes separated from the bulk of the opacity. The nonlinear endpoint of the linear instability is a distribution of 'clouds' of high density and low entropy, idealized here as spheres with some distribution of mass and radius, embedded in a low-density, high-entropy intercloud medium. Characteristic cloud masses and radii are shown in Fig. 1. These characteristics determine what properties the clouds have, and can be compared with the observed statistical properties of intergalactic clouds and galaxies.

The most important scale governing nonlinear behaviour of clouds is the Jeans mass¹⁰, representing a balance of pressure and gravity

$$M_{\text{Jeans}} = RkT/Gm_H \quad (6)$$

is shown in Fig. 1 for a typical final temperature, $T = 10^4 \text{ K}$. The Jeans mass defines the boundary between two types of behaviour for the clouds. On scales below the Jeans mass, self gravity is unimportant and clouds can stabilize close to pressure equilibrium; above the Jeans mass, clouds are unstable to gravitational collapse.

The maximum mass M_{max} of unstable fluctuations is determined by the condition $-\dot{\omega} = H$, related to the Jeans mass by

$$M_{\text{max}} = M_{\text{Jeans}} \left(\frac{E_{e\gamma}}{\gamma kT} \right)^{3/2} \left(\frac{\omega_i}{H} \right)^{3/2} \left(\frac{\bar{\rho}}{\rho_{\text{tot}}} \right)^{3/2} \quad (7)$$

This allows for a possible dark matter contribution to the total density ρ_{tot} , and for the growth rate $-\dot{\omega}$ to exceed the frequency of normal pressure waves $v_s k$ (using equation (4) and $v_s k_{\text{Jeans}} \approx H(\bar{\rho})$).

The lower size limit of the clouds is determined by the photon mean free path $(\bar{x}\bar{n}_H\sigma)^{-1}$, where $\sigma = 6.3 \times 10^{-18} (h\nu/\text{Ryd})^{-3} \text{ cm}^2$ denotes the photoionization cross-section. The characteristic 'shielding mass'

$$M_{\text{shield}} = \frac{4\pi\bar{\rho}}{3(\bar{x}\bar{n}_H\sigma)^3} \quad (8)$$

defines a lower limit to the mass of clouds as a function of their size. Figure 1 shows M_{shield} for a half-neutral medium and $h\nu = 2 \text{ Ryd}$.

The instability produces clouds distributed over a range of initial masses and radii between M_{shield} and M_{max} , in general both above and below M_{Jeans} , along a line of constant density as shown in Fig. 1. Whereas M_{shield} and M_{max} depend just on gas photon physics, the density depends on mean baryon density Ω_b at the epoch of the instability: $n_H = 2.9 \times 10^{-4} \text{ cm}^{-3} [(1+z)/10]^3 (\Omega_b h^2/0.025)$ where $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and z is redshift.

Clouds below the Jeans mass achieve a well-known stable nonlinear configuration, as static density condensations in pressure equilibrium with a hotter, rarefied, more highly ionized intercloud medium. This configuration corresponds qualitatively to the standard model of the quasar Lyman- α absorption clouds^{11,12}. The instability naturally produces such condensations as a byproduct of the ionization of the medium, including a natural explanation (the minimum scale M_{shield}) for their large size¹³. The instability also naturally produces a clean intercloud medium², with high ionization and low density, because the sweeping of gas on small scales happens very quickly.

Clouds above the Jeans mass undergo gravitational collapse and become galaxies. The smallest galaxies should have about the same spatial frequency as the largest Lyman- α clouds, a similar total mass content, similar autocorrelation and a similar characteristic mass. Recent studies of very faint galaxies reveal a population of numerous, faint, relatively unclustered dwarf galaxies which roughly matches this description¹⁴⁻¹⁶, encouraging the idea that the galaxies and intergalactic clouds have a common origin. Because of their dissipative gravitational collapse, collapsed galaxies have a smaller size (hence a smaller sky covering factor) and a higher internal pressure (hence a larger neutral fraction) than the intergalactic clouds, appearing in absorption as the rarer, higher column density, optically thick 'Lyman limit' and 'damped' Lyman absorption systems¹⁷⁻²⁰.

This instability should be considered as a possible additional new ingredient in theories of galaxy and cosmic structure formation. In models including cold dark matter, collisionless dark matter haloes, rather than gas hydrodynamics, both confine the intergalactic clouds and trigger galaxy collapse. If galaxies form because of gas instability, the dark matter halo falls in afterwards²¹, rather than leading the baryonic collapse²². Baryonic galaxies can accrete either hot or cold dark matter; at late times, there is little difference between the two (the typical thermal velocity of a massive neutrino is $v \approx kT_\nu c/m_\nu c^2 \approx 0.8 \text{ km s}^{-1} (60 \text{ eV}/m_\nu)(1+z)$, where $T_\nu = (4/11)^{1/3} T_{\text{cbr}}$ is the temperature of massless neutrinos and T_{cbr} is the temperature of the photon background). If most of the dark matter is neutrinos, this process naturally explains why the phase-space density of Galaxy dark matter roughly coincides with that of primaeval neutrinos²³. It also solves the serious difficulty in neutrino-dominated models of forming galactic-scale structure in spite of the damping of small-scale primordial fluctuations. In principle, it can remove the need for primordial fluctuations altogether. The instability does not however explain large-scale galaxy clustering, which still requires either primordial fluctuations^{24,25} or some separate astrophysical mechanism^{26,27}.

A closure density of massive neutrinos with a very long radiative decay lifetime^{3,4} produces an attractive candidate for the source of both photons and dark matter. For a closure density of one type of neutrinos (say ν_τ), the mass is given by the usual relation $m_\nu = 91.5 h^2 \text{ eV}$; if they decay into a much lighter neutrino and a photon, the photon energy $h\nu = m_\nu c^2/2$ probably exceeds 1 Rydberg. The ionization rate is $\omega_i = n_\tau/n_H \tau_\tau$, where $n_\tau = 10^2 \text{ cm}^{-3}$, yielding for a flat universe

$$\frac{\omega_i}{H} = 0.5 \left(\frac{\tau_\tau}{10^{25} \text{ s}} \right)^{-1} (\Omega_b h^2/0.025)^{-1} h^{-1} \left(\frac{1+z}{10} \right)^{-3/2} \quad (9)$$

Although Sciama³ has cited a variety of observational evidence for neutrinos with lifetimes around 10^{23} s , cosmic ionization (accompanied by instability) occurs sufficiently early even for

lifetimes as large as 10^{25} s. (Ionization constraints in the observed intergalactic medium are satisfied for lifetimes as long as 10^{24} s; ref. 28.) Neutrinos with such a long lifetime would so far have escaped direct detection^{29,30} or any other constraints, and still provide both cosmic ionization and instability. For some parameters, neutrino line radiation could be directly detected from dark haloes at high redshift by the Hubble Space Telescope Faint Object Spectrograph, or from our own galactic halo by the Extreme Ultraviolet Explorer. □

Received 8 April; accepted 17 July 1992.

- Schneider, D. P., Schmidt, M. & Gunn, J. E. *Astr. J.* **101**, 2004–2016 (1991).
- Steidel, C. C. & Sargent, W. L. W. *Astrophys. J.* (in the press).
- Sciama, D. *Comm. Astrophys. Space Sci.* **15**, 71–86 (1990).
- Scott, D., Rees, M. J. & Sciama, D. *Astr. Astrophys.* **250**, 295–301 (1991).
- Fukugita, M. & Kawasaki, M. *Astrophys. J.* **353**, 384–398 (1990).
- Hogan, C. J. & White, S. D. M. *Nature* **321**, 575–578 (1986).
- Hogan, C. J. *Nature* **350**, 469–473 (1991).
- Field, G. B. *Astrophys. J.* **142**, 531–567 (1965).
- Flannery, B. P. & Press, W. H. *Astrophys. J.* **231**, 688–696 (1979).

- Shu, F. H. *The Physics of Astrophysics, Vol. 2 Gas Dynamics* (University Science, Mill Valley, 1992).
- Black, J., *Mon. Not. R. astr. Soc.* **197**, 553 (1981).
- Charlton, J., Salpeter, E. & Hogan, C. J. *Astrophys. J.* (in the press).
- Smette, A. *et al. Astrophys. J.* (in the press).
- Tyson, J. A. *Astr. J.* **96**, 1–23 (1988).
- Cowie, L. L., Lilly, S. J., Gardner, J. & McLean, I. S., *Astrophys. J.* **332**, L29–L32 (1988).
- Efstathiou, G. P., Bernstein, G., Katz, N., Tyson, J. A. & Guhathakurta, P. *Astrophys. J.* (in the press).
- Sargent, W. L. W., Steidel, C. C. & Boksenberg, A. *Astrophys. J.* **69**, 703 (1989).
- Steidel, C. C. *Astrophys. J.* **74**, 37 (1990).
- Wolfe, A. M., Turnshek, D. A., Smith, H. A. & Cohen, R. E. *Astrophys. J.* **61**, 249–304 (1986).
- Lanzetta, K. M. *Astrophys. J.* **375**, 1 (1991).
- Gunn, J. E. *Astrophys. J.* **218**, 592–598 (1977).
- Ryden, B. S. & Gunn, J. E. *Astrophys. J.* **318**, 15–31 (1987).
- Tremaine, S. & Gunn, J. E. *Phys. Rev. Lett.* **42**, 407 (1979).
- Smoot, G. F. *et al. Astrophys. J.* (submitted); *COBE Preprint* 92-02 (1992).
- Wright, E. L. *et al. Astrophys. J.* (submitted); *COBE Preprint* 92-06 (1992).
- Dekel, A. & Rees, M. J. *Nature* **326**, 455 (1987).
- Babul, A. & White, S. D. M. *Mon. Not. R. astr. Soc.* **253**, 31p (1991).
- Dodelson, S. & Jubas, J. M. *Fermilab Preprint* 92/60-A (1992).
- Davidson, A. F. *et al. Nature* **351**, 128 (1991).
- Fabian, A. D., Naylor, T. & Sciama, D. W., *Mon. Not. R. astr. Soc.* **350**, 461 (1991).

ACKNOWLEDGEMENTS. I thank M. J. Rees, who has independently arrived at many of these ideas, for pointing out a basic error in an earlier manuscript, and D. Sciama and E. Bradford for useful comments. This work was supported by NASA.

Dust from short-period comet P/Schwassmann–Wachmann 1 and replenishment of the interplanetary dust cloud

M. Fulle

Osservatorio Astronomico, Via Tiepolo 11, I-34131 Trieste, Italy

IF the current abundance of interplanetary dust is representative of its long-term value¹, there must be a source of dust replenishing the $\sim 10^7 \text{ g s}^{-1}$ (mostly in the form of particles of 10^{-4} to 1 g) destroyed by dissipative processes². Short-period comets are the most likely such source³, but their dust production rate is uncertain. Coma spectrophotometry of several short-period comets excludes significant contributions of dust from them^{4,5}, but observations by the Infrared Astronomical Satellite⁶ have suggested the opposite. Here I use a numerical model⁷ to analyse an optical image of the dust tail of comet P/Schwassmann–Wachmann 1, which contains information about grains $5 \mu\text{m}$ to 2 cm in diameter, ejected from 900 to 30 days before perihelion. During the three years covered by the model, the mass loss rate reached an estimated $(6 \pm 3) \times 10^5 \text{ g s}^{-1}$, for an assumed albedo⁹ of 0.1 at the observation phase angle of 4° . This one short-period comet thus apparently provides $\sim 6\%$ of the mass required to balance the losses of the interplanetary dust cloud.

The peculiar short-period comet P/Schwassmann–Wach-

mann 1 (SW1) is a giant comet¹⁰ orbiting in an almost circular (probably chaotic) path lying between Jupiter and Saturn. Its position may strongly influence its dust loss and may explain the recurrent outbursts for which SW1 is well known¹¹. The rate of dust loss from the SW1 can be used to estimate the contribution made by sources beyond Jupiter to the balance of the interplanetary dust cloud, even if the way in which Jupiter's mass perturbs grains spiralling into the inner Solar System remains unclear. Recently, CCD coma photometry of SW1¹² has indicated a very low rate of dust mass loss, only 10^4 g s^{-1} . This result indicates that a comet beyond Jupiter is too far from the Sun to give a significant dust input to the Solar System, however large its nucleus is.

But the disagreement between the results of photometric and IRAS observations means that we should not consider these results as conclusive¹³. In photometry of the coma, we observe small and large grains overlapping in the field of observation, so that the received radiation is dominated by light of wavelength very close to the size of the dust particles that have reflected the light. Calculations of the dust mass from infrared or visible photometry of the coma therefore contain the implicit assumption that the dust mass is dominated by micrometre-sized or smaller grains; in turn, this implies that the dust size distribution has a power index lower than -4 . In contrast, dynamical analyses of the dust trails detected by IRAS along the orbit of some short-period comets¹⁴ show that they are dominated by large (up to centimetre-sized) grains⁶. This indicates that a size distribution power index lower than -4 is unrealistic, as confirmed by *in situ* measurements for P/Halley¹⁵.

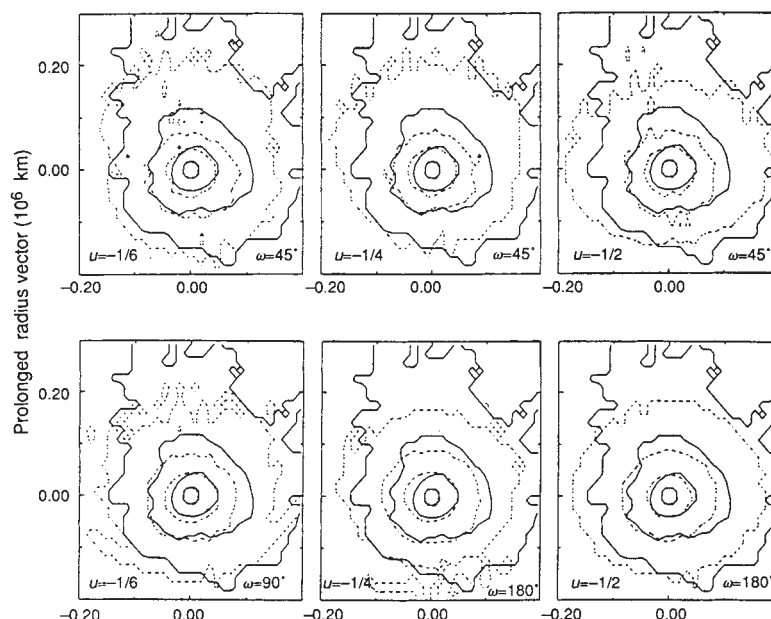
An independent approach to obtaining the dust environment

TABLE 1 Dust tail models of Comet P/Schwassmann–Wachmann 1

ω	u	N_s	N_d	N_t	N'_t	N'_d	T	χ^2	Line style in Figs 2 and 3
180°	-1/6	2578	100	60	15	10	16	1.34	
180°	-1/4	2578	100	60	15	10	14	0.72	short-dashed
180°	-1/2	2578	100	60	15	10	8	0.59	long-dashed
90°	-1/6	1285	100	60	15	10	11	0.84	continuous
90°	-1/4	1285	100	60	15	10	16	0.95	
90°	-1/2	1285	100	60	15	10	11	1.01	
45°	-1/6	382	100	60	15	10	13	0.85	dotted
45°	-1/4	382	100	60	15	10	12	0.71	dot-dashed
45°	-1/2	382	100	60	15	10	12	0.84	dot-dot-dashed

ω , Half-width of the dust ejection cone with axis pointing to the Sun: isotropic ejection, $\omega=180^\circ$; hemispherical ejection, $\omega=90^\circ$; strongly anisotropic ejection, $\omega=45^\circ$. $u=\partial \log v(t, d)/\partial \log d$. N_s, N_d, N_t , dust samples on a dust shell, in size and in time. N'_t, N'_d , samples of the solution vector F in time and in size. T , number of trial functions $v(t, d_0)$. χ^2 , fit quality (integrated over all the 900 image cells) expressed in 10^{-13} mean intensity of the solar disk. The models not shown in Figs 2 and 3 gave unacceptable fits to the input image.

FIG. 1 Isophotes of the dust tail for the intensity levels 0.9, 1.9, 3.5 and 14.0 (10^{-14} mean intensity of the solar disk). The distances along the axes are expressed in 10^6 km units. Continuous lines: observed isophotes. Dashed lines: model isophotes. The image is sampled in 30×30 pixels. ω is the anisotropy parameter, $u = \partial \log v(t, d) / \partial \log d$ (Table 1). The error of the fit of all the isophotes is lower than the noise level (the noise of the background is equal to the intensity level of the outermost isophote).



of comets is given by the photometric analysis and modelling of dust tails¹⁶, which for short-period comets has supplied dust loss rates¹⁷ very close to the values given by IRAS. These analyses allow us to obtain (directly from the data) the dust loss rate and the size distribution⁷; the resulting power index has sometimes been closer to -3 than to -4 . Because centimetre-sized grains can leave the surface of the nucleus and be dragged out by the expanding gas in the coma, the dust masses obtained by coma photometry may be underestimated up to a factor of 10^4 . We can infer reasonable values of the dust mass from comets only if we determine the size distribution of the same grains.

Here we apply the inverse method of analysis⁷ to a CCD image (Fig. 1, passband of 3 nm centred at 642 nm) of the dust tail of comet SW1⁸ obtained on 3 September 1989 with a focal reducer applied to the 2-m Ritchey-Chrétien telescope of the Bulgarian National Observatory. Details of the observations⁸ and of the inverse method⁷ can be found elsewhere. In brief, the model considers $N_t \times N_d \times N_s$ sample dust grains, where N_t , N_d and N_s are the numbers of samples in time t , of grain diameter d , and on a shell uniformly covered by grains of size d ejected from the inner coma at time t with a velocity $v(t, d) = v(t, d_0) \times (d/d_0)^u$. The input image of SW1 contains information on grains of sizes between $5 \mu\text{m}$ and 2 cm .

The solution vector F (which is sampled in $N'_t \times N'_d$ values) is the product of the loss rate and the dust size distribution, and is directly obtained from a least-squares fit to the input image. The inversion of the corresponding linear functional that models the dust tail is unstable and must be regularized by means of a smoothing functional, following the techniques of ill-posed inverse problems¹⁸. The inverse relation between the dust sizes and the ratio of solar radiation pressure to gravity depends linearly on the dust scattering efficiency Q_{rp} and inversely on the dust bulk density, for which we assume a reference value of 1 g cm^{-3} . The sizes must be scaled inversely to other densities^{7,19}. In contrast the dust mass calculated by means of dust tail analysis is independent of the dust density¹⁶, whereas it depends linearly on Q_{rp} and inversely on the albedo⁷, the upper limit of which is 0.1 (ref. 9) for the observation phase angle of 4° .

The scattering efficiency is given by $Q_{rp} = Q_{abs} + (1 - g)Q_{sca}$, which, for forward-throwing conditions ($g \approx 1$, where g is the mean cosine of the scattering angle) and for a very low grain albedo, becomes $Q_{rp} \approx Q_{abs}$. The grains considered by the model have sizes much larger than the wavelength of the solar light, so a good approximation is the geometrical limit $Q_{abs} = 1$. This

conclusion holds also for fluffy grains, as they will still have indices of refraction such that $Q_{rp} \approx Q_{abs}$ for the large sizes considered. Other nonlinear free parameters are the dust ejection anisotropy ω , the velocity power index u and the time dependence of the ejection velocity $v(t, d_0)$ for a diameter $d_0 = 0.03 \text{ cm}$. The tail can be fitted only with a well defined velocity, which can be found by testing many trial functions (Table 1), and therefore it is a result and not an assumption of the model.

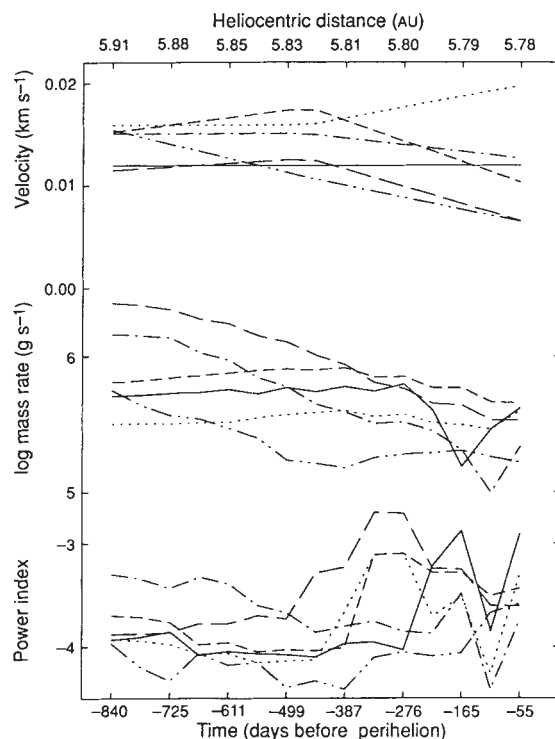


FIG. 2 Dust environment of comet P/Schwassmann-Wachmann 1: the dust ejection velocity $v(t, d_0)$ (for a diameter $d_0 = 0.03 \text{ cm}$ if the dust bulk density is 1 g cm^{-3} , $d_0 = 0.3 \text{ cm}$ if the density is 0.1 g cm^{-3}), the dust mass loss rate (assumed $Q_{rp} = 1$ and albedo of 0.1 for a phase angle of 4° , the loss rate depends linearly on Q_{rp} and inversely on the albedo), the power index of the time-dependent size distribution. The relation between line style and parameter combination is given in Table 1.

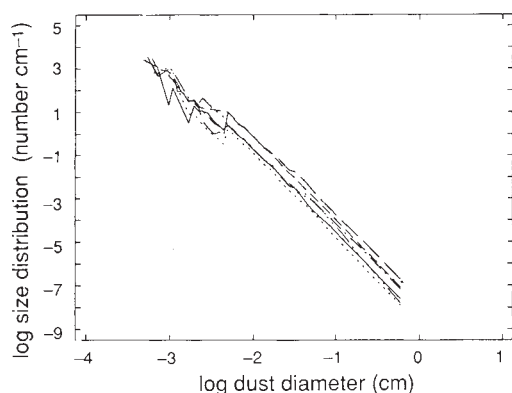


FIG. 3 Time-averaged and normalized size distribution for a dust bulk density of 1 g cm^{-3} (the diameters depend inversely on the assumed density). See Table 1 for line styles. The power index of the distribution of P/Schwassmann-Wachmann 1 is -3.3 ± 0.3 .

The results are plotted in Fig. 2, and in Fig. 1 the corresponding fits of the observed tail are reconstructed by means of the solution F . The index $u = -\frac{1}{2}$ gives velocities and loss rates decreasing in time; these are unrealistic results which should be rejected because the comet approaches the Sun. The instabilities of the loss rate for $t > -200$ may be an artefact due to the inversion of the ill-posed problem²⁰. If the size distribution has a power index larger than -4 , the dust mass is independent of the lower limit of the sizes but strongly dependent on the upper one. The ejection of grains much larger than 2 cm at 6 AU from the Sun is, however, improbable. The most reliable error estimate of the mass loss rate is given by the dispersion of the results for different combinations of the free parameters u and w (ref. 20).

Our results give a mass loss rate of $(6 \pm 3) \times 10^5 \text{ g s}^{-1}$. Because all the grains are injected in bound orbits, the same rate of change of mass goes to replenish the interplanetary dust cloud. This contribution alone balances $6 \pm 3\%$ of the required mass in the interplanetary cloud². The rate of loss may be too large to be explained by water-ice sublimation¹²; phase transitions between water-ice states²¹ may account for the outbursts which probably eject much smaller particles²² and therefore lower dust masses. The different conditions of dust ejection and the uncertainties in models of gas drag on very large grains²³ do not allow conclusive comparisons between our velocities and the lower ones derived from analysis of the P/Tempel 2 trail¹⁴. Extended observations¹² combined with our mass loss rate, explain the persistent coma as the result of steady activity in SW1 rather than from overlaps of outburst shells²⁴.

The high mass-loss rate is consistent with the power index of the time-averaged size distribution (-3.3 ± 0.3 , Fig. 3). This power index is higher than the value obtained for comets P/Encke and P/D'Arrest¹⁷, and is close to the index of the dust released by the new comet Wilson 1987VII well before perihelion²⁵ ('new' in the sense that it has recently entered the Solar System from the Oort cloud). Because the dust of P/Encke and P/D'Arrest was released within the orbit of Mars, such differences suggest that the size distribution of comet dust depends on its distance from the Sun when it is produced, not only on the comet's age. A dust production process different from the usual one (which is based on a water-dominated coma²⁶) is therefore plausible and compatible with the large mass loss rate and the observation of a CO^+ tail⁸. The results suggest a coma dominated by CO or CO_2 beyond Mars. □

4. Newburn, R. L. & Spinrad, H. *Astr. J.* **90**, 2591–2608 (1985).
5. Newburn, R. L. & Spinrad, H. *Astr. J.* **97**, 552–565 (1989).
6. Sykes, M. V. in *Origin and Evolution of Interplanetary Dust* (Reidel, Dordrecht, in the press).
7. Fulle, M. *Astr. Astrophys.* **217**, 283–297 (1989).
8. Jockers, K., Bonev, T., Ivanova, V. & Rauer, H. *Astr. Astrophys.* **260**, 455–464 (1992).
9. Hanner, M. S. & Newburn, R. L. *Astr. J.* **97**, 254–261 (1989).
10. Rickman, H. in *Dynamics of Comets: Their Origin and Evolution* (ed Carusi, A. & Valsecchi, G. B.) 149–172 (Reidel, Dordrecht, 1985).
11. Froeschle, C., Klinger, J. & Rickman, H. in *Asteroids, Comets, Meteors* (eds Lagerkvist, C. & Rickman, H.) 215–224 (Uppsala Univ. Press, Uppsala, 1983).
12. Jewitt, D. *Astrophys. J.* **351**, 277–286 (1990).
13. Delsemme, A. H. in *Diversity and Similarity of Comets* ESA SP-278, 19–30 (1987).
14. Sykes, M. V., Lien, D. J. & Walker, R. G. *Icarus* **86**, 236–247 (1990).
15. McDonnell, J. A. M., Lamy, P. L. & Pankiewicz, G. S. in *Comets in the Post-Halley Era* (eds Newburn, R. L. Jr, Neugebauer, M. Rahe, J.) 1043–1074 (Kluwer, Dordrecht, 1991).
16. Finson, M. L. & Probst, R. F. *Astrophys. J.* **154**, 327–352 (1968).
17. Fulle, M. *Astr. Astrophys.* **230**, 220–226 (1990).
18. Marquardt, D. W. *Technometrics* **12**, 591–602 (1970).
19. Burns, J. A., Lamy, P. L. & Soter, S. *Icarus* **40**, 1–48 (1979).
20. Fulle, M., Cremonese, G., Jockers, K. & Rauer, H. *Astr. Astrophys.* **253**, 615–624 (1992).
21. Smoluchowski, R. in *Ices in the Solar System* (eds Klinger, J., Benest, D., Dollfus, A. & Smoluchowski, R.) 397–406 (Reidel, Dordrecht, 1985).
22. Cochran, A. L., Cochran, W. D. & Barker, E. S. *Astrophys. J.* **254**, 816–822 (1982).
23. Crifo, J. F. in *Comets in the Post-Halley Era* (eds Newburn, R. L. Jr, Neugebauer, M. & Rahe, J.) 937–990 (Kluwer, Dordrecht, 1991).
24. Sekanina, Z. *Astr. J.* **100**, 1293–1314 (1990).
25. Cremonese, G. & Fulle, M. *Astr. J.* **100**, 1285–1292 (1990).
26. Delsemme, A. H. in *Comets* (ed Wilkening, L. L.) 85–130 (Univ. of Arizona Press, Tucson, 1982).

Coalescence reactions of fullerenes

Chahan Yeretzian*, Klavs Hansen*†, François Diederich‡ & Robert L. Whetten*

* Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024-1569, USA

† Niels Bohr Institute, 4000 Roskilde, Denmark

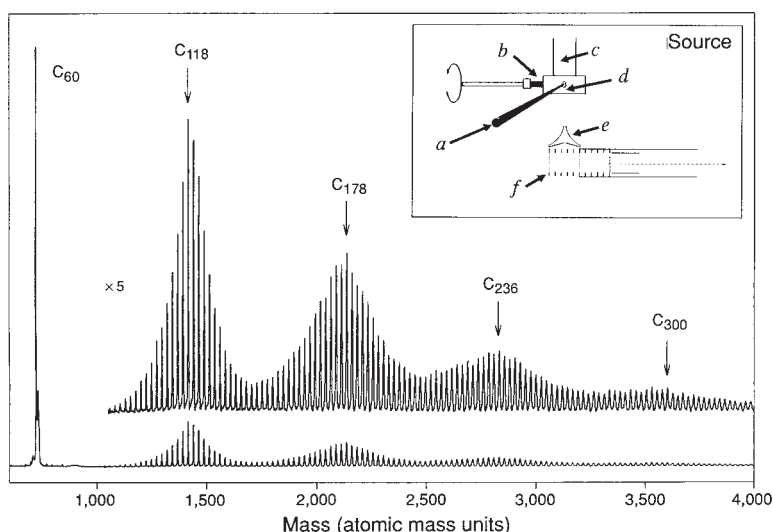
‡ Laboratorium für Organische Chemie, ETH-Zentrum, Universitätsstrasse 16, CH-8092 Zürich, Switzerland

THE production of fullerene molecules typically involves extreme high-temperature conditions (electric arcs¹, flames² or resistive heating³), and the reactive processes involved are poorly understood. Once separated^{4,5}, these molecules can undergo several important reactions, including formation of charge-transfer^{6,7} and adduct^{8,9} compounds, and the encapsulation of atoms^{10–12}. Here we present evidence for coalescence reactions between fullerene molecules: mass spectrometric measurements on hot, dense vapours of small fullerenes (C_{60} and C_{70}) reveal the formation of stable higher fullerenes which are multiples of the initial masses. The heat of coalescence is released through emission of small, even-numbered fragments which, in a very dense vapour, are efficiently captured by other coalesced fullerenes. These findings have implications for the mechanisms of fullerene formation and growth, and may open the way to new synthetic routes to selected higher fullerenes and encapsulation compounds.

The fullerenes are a series of highly cohesive all-carbon molecules, C_{2p+20} , whose structures are closed bonding nets consisting of 12 pentagonal and p hexagonal rings^{13,14}. Giant fullerenes, with very large p , may therefore have properties tending toward those of a single, closed graphitic sheet: an infinite, or periodic, extension of six-membered rings. For the molecule C_{60} , the cohesive energy per atom is already only 0.4 eV per atom below (less stable than) graphite, recovering 95% of the bulk binding^{15,16}. First-principles and semi-empirical calculations^{17–19} indicate that the cohesion per atom increases steadily with p , from C_{60} to C_{70} and towards C_{96} ; a particularly stable form of C_{120} is estimated to be 0.16 eV per atom, or nearly 20 eV, more stable than two C_{60} molecules^{17,18}. This increase in cohesive energy per atom provides a strong driving force for fullerene coalescence: reactions in which two small fullerenes merge and open up to form a larger net or giant fullerene. Given the clear existence of this thermodynamic driving force, the only question is one of the kinetics of the growth at realistic temperatures and densities.

Received 9 December 1991; accepted 14 July 1992.

1. Maurette, M., Jehanno, C., Robin, E. & Hammer, C. *Nature* **328**, 699–702 (1987).
2. Grün, E., Zook, H. A., Fechtig, H. & Giese, R. H. *Icarus* **62**, 244–272 (1985).
3. Whipple, F. L. in *Zodiacal Light and the Interplanetary Medium* (ed. Weinberg, J.) NASA SP-150, 409–426 (1967).



We have observed these processes using the following experimental approach (Fig. 1). A pulse of hot, dense vapour is generated by laser desorption of a fullerene film into an inert gas flowing at near-sonic speed. After a short time ($<10^{-3}$ s), the gas expands into vacuum, terminating molecular collisions, and the composition of the gas mixture is measured by mass spectrometry (for details, see Fig. 1). Mass spectra are signal-averaged over many such pulses, and it has been verified that the purified fullerene film is not transformed by this treatment. Substantial quantities of material are consumed, of the order of 1 mg per hour.

Under ordinary low-yield conditions (low fluence, small irradiated spot size or no carrier gas), the desorbed vapour consists entirely of the monomeric species present in the solid, with essentially no fragmentation⁴. Desorption with a tightly focused laser gives the monomer and its fragments. Use of large-area irradiation at higher total intensities can yield mass distributions showing negligible fragmentation to smaller fullerenes, but with a very broad, nearly smooth distribution of C_{2n} species extending from C_{70} to C_{400} or higher (not shown), which is not readily distinguished from those produced by, for example condensation of laser-generated graphite vapour¹³.

Through careful control of the vapour density and heating, achieved by varying the laser fluence and irradiated spot size, we can obtain very different distributions of giant carbon molecules (Figs 1 and 2). For C_{60} vapour in Fig. 1, the distribution varies smoothly, but has strong oscillations with peaks near multiples of C_{60} : near C_{120} , C_{180} , C_{240} and C_{300} . Figure 2a shows a similar pattern at slightly higher fluences, where a greater fraction of the C_{60} is converted to higher masses. This pattern indicates the aggregation of individual molecules into giant aggregates, $mC_{60} \rightarrow C_{60m}$, although other processes are clearly involved.

To clarify the processes leading to distributions such as those in Figs 1 and 2a, the extent of reaction must be much more restricted. We did this by decreasing the fluence and increasing the sensitivity to the few high-mass species formed (Figs 2b, 2c and 3). The mass distribution about each maximum becomes asymmetric, with step-like decreases displaced from the simple combination product by a single C_2 unit: twice C_{60} yields predominantly C_{118} , twice C_{70} gives C_{138} , and C_{60} plus C_{70} gives C_{128} . Accompanying each main peak is a well developed series corresponding to a stream of losses; for example, twice C_{70} yields C_{138} to C_{120} in a sharply truncated distribution. When no carrier gas is present to confine the desorbed vapour, these processes are barely observable, and give a very asymmetric distribution (Fig. 3, upper panel).

FIG. 1 Distribution of carbon clusters produced by high-intensity irradiation of pure C_{60} powder under a helium flow (detecting positively charged clusters). All the higher peaks are even-numbered. To the far left is the solitary C_{60} peak. The maxima indicated are near-multiples of the C_{60} starting material, that is, near C_{60p} where $p=2, 3, 4, 5$. Note the nearly complete absence of the fragments $n=58$ and 56 , and also the next higher special fullerenes, $n=70, 84$ and so on. Inset, detail of the source. To generate the spectrum shown, a slightly focused ultraviolet laser pulse *a* (Nd:YAG, 266 nm, 6 ns) at a fluence near 2 mJ cm^{-2} (at least three times as high as used in conventional compositional analysis⁴), is fired at the sample *b* simultaneously with a 10-bar helium flow from a pulsed gas valve *c*. Desorbed molecules, both charged (positive and negative) and neutral, are swept along by the helium jet through a flow channel *d*, 2 mm long by 4 mm wide and a skimmer *e*, and into the pulsed extraction region *f* of a reflectron time-of-flight spectrometer. Charged clusters are extracted perpendicular to the helium flow by an electric-field pulse, so that the charged component of the vapour is detected without further exposure to ionizing radiation.

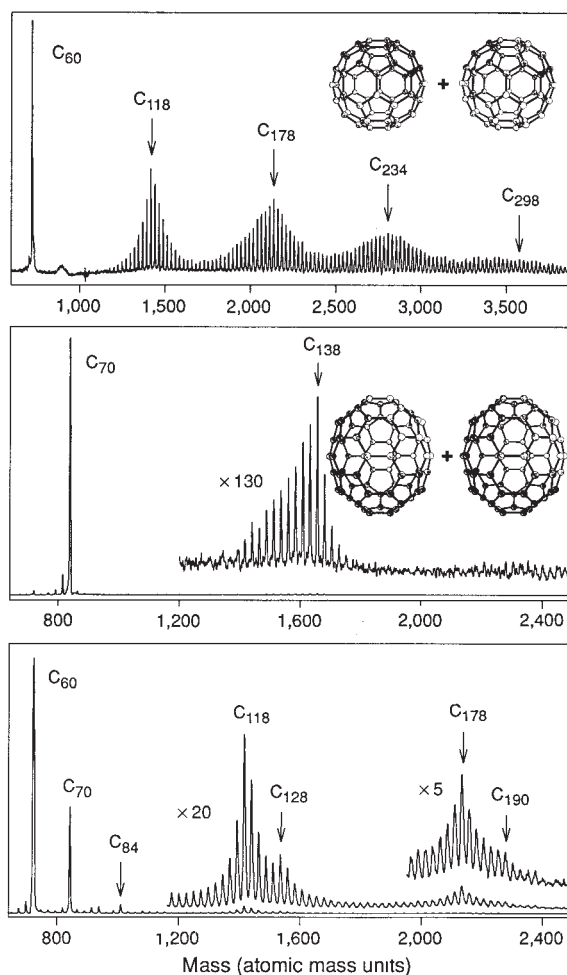


FIG. 2 Distributions generated as in Fig. 1, but for the specific fullerene materials and conditions as follows. *a*, Pure C_{60} sample, at higher laser fluence than that used in Fig. 1. *b*, Pure C_{70} sample, at very low fluence. The peak at C_{138} is also a ledge. The fragmentation extending downwards from C_{138} goes as far as C_{118} or even lower. *c*, Toluene extract of Krättschmer-Huffman soot, containing soluble fullerenes C_n , where $n=60, 70, 76, 78, 82, 84$, and trace amounts of higher fullerenes. Local maxima appear at $n=118, 128, 178, 190$, corresponding to combinations of nC_{60} and nC_{70} . (n, m) = (2, 0), (1, 1), (3, 0), (2, 1).

To verify that the reaction products are indeed strongly bound, rather than being simple aggregates held together by polarization forces or singly linking bonds, we collided the mass-selected C_{118}^+ beam with a solid surface (Si(111) terminated by hydrogen) at high velocity¹⁹, and looked at the fragmentation pattern. As shown in Fig. 4, only a strongly scattered (intact) parent ion (C_{118}^+) and no fragments could be detected at energies of at least 200 eV. This result is similar to that obtained on other fullerenes^{19–21}. Fullerene adducts, bound by pairs of external bonds, are readily broken at much lower energies (C.Y. and R.L.W., manuscript in preparation). A lower bound of ~ 6 eV on the energy of processes such as $C_{118} \rightarrow C_{60} + C_{58}$ is thereby estimated, using standard statistical rate theory arguments and energy-transfer assumptions documented elsewhere^{19–22}. This value is inconsistent with a small number of bonds linking the cages, but is consistent with the integrity of known fullerene molecules. Furthermore, the shift in the time-of-flight indicates that these collisions are highly elastic up to an energy of 80 eV, and are increasingly inelastic thereafter.

Because coalescence products are strongly enhanced by the presence of a confining external gas (helium), it is clear that the processes are occurring in the dense vapour phase on a timescale of $\sim 10^{-5}$ s, rather than in the solid on a timescale of 10^{-7} s. This vapour is estimated to contain typically 10^{13} molecules in a volume of typically 10^{-3} cm³, so that fullerene-fullerene collisions initially occur with very high frequency. The almost complete absence of smaller fragments such as C_{58} and C_{56} in all these distributions argues against a significant role for them in this process. The vapour temperature is not easily estimated, but the absence of C_{60} fragments, estimated to require 11 eV (ref. 16), gives an upper bound.

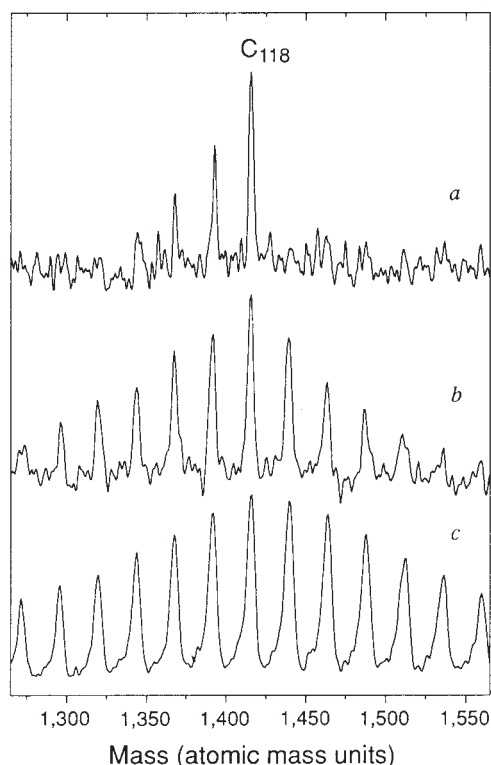


FIG. 3 Fluence dependence and effect of carrier gas on the coalescence product distribution obtained from C_{60} , shown in the region C_{106} to C_{130} . In *a*, no carrier gas was used, so there was no confinement of the desorbed vapour. In *b* and *c*, the carrier gas was present, but the fluence was reduced strongly in *b* from the conditions in Figs 1 and 2*a*. Decreasing the fluence or removing the helium carrier gas both cause a transition from a symmetric distribution around C_{118} (*c*), for higher vapour density, to an asymmetric one with a ledge at C_{118} (*b*, *a*) at lower vapour density. At the low-density conditions of *a*, the direct coalescence product C_{120} is completely absent, as are products of capture processes (C_{122} and higher).

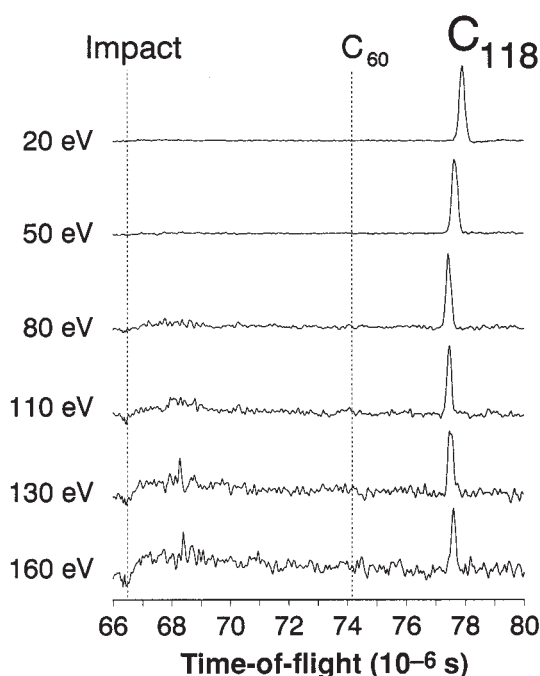


FIG. 4 Impact of C_{118}^+ , mass-selected from a coalescence distribution, against a Si(111) surface. These time-of-flight mass spectra of scattered clusters show a complete absence of fragmentation, for example to C_{60} plus C_{58} , for energies to at least 160 eV. The approximate locations of the time of impact and the time where the C_{60} fragment would appear are marked. The time-shift of the scattered C_{118}^+ signal reflects the elasticity of the collision at low energy (shift to earlier time) and inelasticity at higher impact energy (shift toward later time). By comparison to known processes (such as $Na_{63}F_{62}$ with 125 atoms), we place a lower bound of 6 eV on the energy of breakup.

The asymmetric mass distributions around each coalescence maximum at the lower fluences indicate that coalescence processes are highly energetic and release excess energy through fragmentation processes, for example $C_n + C_m \rightarrow C_{n+m} \rightarrow C_{n+m-2} + C_2$. Fragmentation by emission of small even-numbered units is characteristic of hot giant fullerenes²³. According to the calculations of Adams *et al.*¹⁸, the rapid increase of the cohesive energy (with size) toward graphite implies an energy release per atom (total) of 0.16 eV (19 eV) for $2 C_{60} \rightarrow C_{120}$, 0.24 eV (43 eV) to produce C_{180} , and 0.28 eV per atom (67 eV) to produce C_{240} . In the hot vapour, this entire energy might be released through emission of two or more small (C_2) or larger (C_4 or C_6) even-numbered fragments. The statistical nature of the cooling is expressed by a large range of fragments, peaking at a single C_2 loss but extending all the way to 16 C loss (for C_{60} dimerization) or 22 C loss (for C_{70} dimerization).

A second effect, appearing as a mass gain that makes the distributions symmetrical under higher-density conditions (Figs 1 and 2*a*), requires that these small emission fragments are captured by previously coalesced fullerenes in the dense vapour to produce heavier species, for example $C_{118} + C_4 \rightarrow C_{122}$. Recalling that the vapour density is very high at the higher fluences, it is natural to propose that the small fragments generated through evaporative cooling do not escape the hot zone, as they would in low-density vapours, but instead collide and react with other fullerenes, including the nascent coalescence products. It may be significant that these capture processes do not occur for C_{60} to C_{70} (note the absence of C_{62} , C_{64} or C_{72}), a fact which may be related to their postulated reluctance to grow in contact with carbon vapour²⁵. The combined evaporation-and-capture reaction process can be regarded as an exchange mechanism

that broadens (more or less symmetrically) the abundance distributions. A detailed mathematical analysis of these distributions, within a kinetic model of coalescence and capture processes, will be presented elsewhere (K.H., manuscript in preparation). An alternative hypothesis for forming species such as C_{122} , involving fullerenes such as C_{62} and C_{64} , is untenable, as they are not present in the vapour.

The experiments described here are of limited use in determining the precise kinetics of coalescence at known temperatures; a precision measurement of the reaction yield requires collecting and weighing the higher fullerene material generated. Experiments towards these goals are in progress. One concern is the role of charge, as we have not detected coalescence products in the negative-ion channel. It is possible that coalescence is accelerated in a plasma environment. (For example, if the initial temperature were 2,500 K, and given the known difference of ~ 4.8 eV between ionization potential and electron affinity for either C_{60} or C_{70} , then the fraction of charged pairs is 10^{-8} , or a total of 10^5 pairs per pulse, in our case.) Further information on the reaction dynamics could be provided by a study of nonthermal processes. Campbell *et al.*²⁴ have collided an accelerated beam (850 eV) of C_{60}^+ ions with thermal C_{60} molecules and found that heavier ions (not mass-resolved, but covering the C_{90} – C_{120} range) are generated.

The existence of efficient coalescence reactions has implications for fullerene research and applications, most obviously concerning the long-time stability of fullerene vapours and plasmas, and for the growth and formation processes, where coalescence must be resisted so as to stabilize C_{60} and C_{70} in high yield. Controlled coalescence might be used to produce large quantities of selected higher fullerenes such as C_{120} , C_{130} and C_{140} , where higher inert gas pressure is required to remove the heat of coalescence. We have begun to investigate this possibility. Coalescence processes may also be involved in the formation of larger encapsulation compounds which can involve the fusion of empty fullerenes in the presence of metal^{26–28}. A related process may be involved in the degradation of pure fullerene solids. For example, Wang and Buseck²⁹ have reported

that electron-beam irradiation in an electron microscope stimulates growth of larger structures that may be fullerenes. □

Received 19 June; accepted 5 August 1992.

1. Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354 (1990).
2. Peters, G. & Jansen, M. *Angew. Chem. Int. Ed. Engl.* **31**, 223 (1992).
3. Howard, J. B., McKinnon, J. T., Makarovskiy, Y., Lafleur, A. & Johnson, M. E. *Nature* **352**, 139–141 (1991).
4. Ajie, H. *et al.* *J. phys. Chem.* **94**, 8630–8633 (1990).
5. Taylor, R., Hare, J. P., Abdul-Sada, A. K., Kroto, H. W. *J. chem. Soc. Chem. Commun.* 1423–1425 (1990).
6. Allemand, P.-M. *et al.* *J. Am. chem. Soc.* **112**, 1050–1051 (1990).
7. Haddon, R. C. *et al.* *Nature* **350**, 320–323 (1991).
8. Suzuki, T., Li, Q., Khemani, K. C., Wudl, F. & Almarsson, Ö. *Science* **254**, 1186–1187 (1991).
9. Krusic, P. J., Wasserman, E., Keizer, P. N., Morton, J. R. & Preston, K. F. *Science* **254**, 1183–1185 (1991).
10. Chai, Y. *et al.* *J. phys. Chem.* **95**, 7564–7568 (1991).
11. Weiske, T., Böhme, D. K., Hrusak, J., Krätschmer, W. & Schwarz, H. *Angew. Chem. Int. Ed. Engl.* **30**, 884–886 (1991).
12. Ross, M. M. & Callahan, J. H. *J. phys. Chem.* **95**, 5720–5723 (1991).
13. Kroto, H. W. *Angew. Chem. Int. Ed. Engl.* **31**, 111–129 (1992).
14. Diederich, F. & Whetten, R. L. *Acc. chem. Res.* **25**, 119–126 (1992).
15. Beckhaus, H.-D., Rüchardt, C., Kao, M., Diederich, F. & Foote, C. S. *Angew. Chem. Int. Ed. Engl.* **31**, 63–64 (1992).
16. Stanton, R. E. *J. phys. Chem.* **96**, 111–118 (1992).
17. Zhang, B. L., Wang, C. A., Ho, K. M. *Chem. Phys. Lett.* **193**, 225–230 (1992).
18. Adams, G. B., Sankey, O. F., Page, J. B., O'Keeffe, M. & Drabold, D. A. *Science* **256**, 1792–1795 (1992).
19. Beck, R. D., St. John, P., Alvarez, M. M., Diederich, F. & Whetten, R. L. *J. phys. Chem.* **95**, 8402–8409 (1991).
20. Busmann, H.-G., Lill, Th. & Hertel, I. V. *Chem. Phys. Lett.* **187**, 459–465 (1990).
21. Busmann, H.-G., Lill, Th., Reif, B. & Hertel, I. V. *Surf. Sci.* **272**, 146–153 (1992).
22. Yeretizian, C. & Whetten, R. L. *Z. Phys. D* (in the press).
23. Maruyama, S., Lee, M. Y., Hauler, R. E., Chai, Y. & Smalley, R. E. *Z. Phys. D*, **19**, 409 (1991).
24. Campbell, E. E. B., Hielscher, A., Ehlich, R., Schjya, V. & Hertel, I. V. in *Nuclear Physics Concepts in the Study of Atomic Cluster Physics* (eds Schmidt, R., Lutz, H. O. & Dreizler, R.) Lecture Notes in Physics Vol. 404, 185 (Springer, 1992).
25. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F., Smalley, R. E. *Nature* **318**, 162–163 (1985).
26. McElvany, S. W., Ross, M. M. & Callahan, J. H. *Acc. chem. Res.* **25**, 162–168 (1992).
27. Jin, C., Guo, T., Chai, Y., Lee, A. & Smalley, R. E. *Fullerene Nanowires, Proc. 1st Italian Workshop on Fullerenes, 6–7 February 1992* (eds Taliani, C., Ruani, G. & Zamboni, R.) (World Scientific, 1992).
28. Smalley, R. E. in *Naval Res. Rev.* **3**, 3–14 (1991).
29. Wang, S. & Buseck, P. R. *Chem. Phys. Lett.* **182**, 1–4 (1991).

ACKNOWLEDGEMENTS. We acknowledge L. Issacs for supplying pure C_{60} and C_{70} , and M. M. Alvarez and R. D. Beck for their roles in early experiments leading to this research. This research has been supported by the NSF, an ONR contract, a postdoctoral fellowship to C.Y. from the Schweizerischer Nationalfonds, a grant to K.H. from the Nørgård Foundation and a Packard Foundation Fellowship to R.L.W.

Enhancements in biologically effective ultraviolet radiation following volcanic eruptions

A. M. Vogelmann, T. P. Ackerman & R. P. Turco*

Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

* Department of Atmospheric Sciences, University of California, Los Angeles, California 90024, USA

AEROSOLS injected into the stratosphere by large volcanic eruptions may induce ozone destruction through processes including heterogeneous chemical reactions. The effect of ozone reductions on surface ultraviolet irradiation is not obvious, however, because aerosols also increase the reflection of sunlight. Here we use a radiative transfer model to estimate the changes in biologically effective ultraviolet radiation (UV-BE) at the Earth's surface produced by the El Chichón (1982) and Mount Pinatubo (1991) eruptions. We find that in both cases surface UV-BE intensity can increase because the effect of ozone depletion outweighs the increased scattering.

Ozone is the primary absorber of the Sun's ultraviolet radiation, which is potentially harmful to plants and animals. It is possible that short-term ozone depletions are caused by large volcanic eruptions, which inject millions of tons (Mt) of SO_2 into the stratosphere where it then forms an H_2SO_4 (sulphuric acid) aerosol. This aerosol provides surfaces on which hetero-

geneous chemical reactions are able to activate chlorine into forms that catalyse ozone destruction^{1,2}. The potential for ozone destruction increases with volcanic aerosol loading and with stratospheric chlorine concentrations. (The primary source of stratospheric chlorine is anthropogenic, as the amount of chlorine injected into the stratosphere by volcanoes seems to be small³.) If the sulphate aerosol loading were to reach certain critical values, it is possible that catastrophic ozone loss rates could occur⁴, similar to those in the Antarctic spring. Further, the aerosol alters the stratospheric radiative and temperature fields, so it may affect photodissociation rate constants so as to decrease ozone concentrations⁵. Although theory suggests that ozone depletions may result from volcanic eruptions, observational evidence for this is uncertain.

El Chichón erupted in March and April of 1982, causing chemical and radiative perturbations on a global scale⁶. Ozone column depletions of 1–7% were observed in many regions of the world, with even larger depletions observed at specific monitoring stations^{7–10}. The ozone depletions typically started in late autumn, peaked in winter, and returned to normal levels by summer. During the December–April period after the eruption, some Northern Hemisphere stations reported the lowest monthly values measured in 40 years^{7,8}. The relationship between these ozone depletions and the volcanic eruption is obscured by other phenomena occurring at that time, such as the westerly phase of the quasibiennial oscillation (QBO), or the abnormally intense El Niño/Southern Oscillation (ENSO)^{7,8}; analyses indicate, however, that the aerosol was responsible for a depletion of 2–10% in the affected layers^{1,8,10}.

Mount Pinatubo erupted in June of 1991, injecting almost

three times as much sulphur into the stratosphere as did El Chichón¹¹; it was thus the largest eruption of this century. Preliminary estimates of the aerosol-induced, zonal average ozone depletions give upper limits of 2–4% (S. Chandra, personal communication). In November, however, an anomalous, regional-scale depletion of 20% from the climatological mean was observed over Boulder, Colorado (D. J. Hofmann, personal communication). These observations suggest that volcano-related ozone deficits could be occurring, as may have followed the El Chichón eruption.

Although the relation is still controversial, observed ozone depletions associated with major eruptions are sufficient for us to analyse the enhanced UV-BE at the Earth's surface. On the basis of observed aerosol loadings and ozone depletions following the El Chichón and Pinatubo eruptions, we use a radiative transfer model to calculate changes in the amount of UV-BE reaching the surface, and from this assess the possible impact of large eruptions on the global UV-BE budget.

The combined effects of ozone depletion and aerosol loading on the transmission of ultraviolet radiation are calculated using a 84-stream, discrete-ordinates-method radiative transfer code¹², in which the fractional error of the computation is less than 1–2% (ref. 13). Ultraviolet fluxes are calculated for a horizontal surface at sea level and include effects of the atmosphere's sphericity, the time-dependent solar geometry, Rayleigh scattering¹⁴, ozone absorption^{15,16}, aerosol scattering and absorption, and surface reflectivity. The solar spectral irradiance at the top of the atmosphere is obtained from the World Meteorological Organization (WMO) study¹⁶. A background tropospheric aerosol profile¹⁷ is used where the aerosol spectral scattering properties are calculated using a rural aerosol model from 0.0 to 2.5 km, and a tropospheric aerosol model from 2.5 to 11.5 km (ref. 18). Aerosol scattering phase functions are represented by a Henyey–Greenstein function. This radiative transfer model produces good agreement with data recorded at Antarctica^{19,20} over a wide range of wavelengths and solar zenith angles.

An increase in ultraviolet radiation, particularly at UV-B wavelengths (290–320 nm), may affect an organism's growth, health, reproductive ability or survival²¹. Biological action spectra are used to approximate the wavelength-dependent sensitivity of an organism's specific biological response when exposed to radiation. Here we have used action spectra corresponding to a DNA damage spectrum²², a generalized plant

damage spectrum²³, and an erythema response²⁴ (sunburn). Also included is the response spectrum for a Robertson–Berger meter²⁵, which is widely used to monitor the environmental levels of ultraviolet radiation. The daily dose of biologically effective ultraviolet energy (DUV-BE) is determined by integrating the product of the daily energy dose at the surface (DUV) weighted by the value of the action spectra at the respective wavelength across the spectral region from 280 to 420 nm in 0.1-nm steps. DUV-BE is a strong function of the latitude, time of year²⁶ and cloud cover²⁷; values for cloudless skies at the Equator are the greatest and, at higher latitudes, summertime values are several times greater than wintertime doses.

The optical properties of volcanic aerosols were calculated from Mie theory. Calculations assumed a log-normal distribution of spherical particles composed of 75% H₂SO₄ by weight (25% water)¹, and volume mode radii ranged from 0.3 to 1.8 μ m, in agreement with observations following the El Chichón and Pinatubo eruptions^{6,28}. On the basis of these computations, the wavelength-invariant aerosol optical properties adopted are a single scattering albedo of 0.99 and an asymmetry parameter of 0.75. Optical depths range from 0.0 to 0.4, bounding zonal-mean values observed after the eruptions^{6,28}. Ozone depletions and aerosol loadings are assumed to occur between the altitudes of 12 and 24 km, in general agreement with observations^{7,8,29}. Modelled ozone depletions and aerosol loadings are volumetrically uniform in the 12–24 km layer; computations are representative of those for observed profiles because the ultraviolet transmission depends primarily on the total amount of ozone and volcanic aerosol present, not on details of the stratospheric distributions.

Our baseline case assumes a United States Standard Atmosphere profile for mid-latitude winter³⁰ with the ozone profile normalized to 350 Dobson units, a snow-free ultraviolet surface albedo of 5% (ref. 31) and a mid-March insolation at 40° N. Results are presented as percentage differences in DUV-BE (between perturbed and unperturbed states) from the baseline case. These results can be considered roughly equivalent to results at other latitudes with different ozone columns whenever the noontime solar zenith angle is less than ~50°; this angle is not exceeded at noon any time between 30° S and 30° N, or in the spring and summer for latitudes less than 50° and 70°, respectively. The effects of water clouds are not explicitly included as their ultraviolet optical properties are wavelength-invariant and percentage differences are very similar to those for cloudless skies.

The increase in DUV-BE for a given ozone depletion depends on the action spectrum considered. This difference is shown in Fig. 1 where, in the absence of a volcanic aerosol, percentage increases in DUV-BE are plotted as a function of ozone column depletion ranging from 0 to 30%; also shown are changes in the daily dose of (non-action-spectra-weighted) spectrally integrated UV-B energy (wavelengths from 290 to 320 nm). All curves show the well known relationship between ozone reduction and ultraviolet dose increase, whereby a 1% change in ozone amount produces an increase of more than 1% in ultraviolet dose. The Robertson–Berger meter response is the least sensitive of the weighted responses to ozone changes; the present ultraviolet monitoring network may therefore underestimate the biological damage resulting from ozone decreases.

DUV-BE changes for ozone depletions concurrent with volcanic aerosol loadings are illustrated in Fig. 2. Only the differences in DUV-DNA (daily dose of ultraviolet energy weighted by DNA action spectrum) are presented, as a given aerosol loading changes all DUV-BE similarly (although the aerosol's effect on transmission can be very different for individual ultraviolet wavelengths³²). Hemispheric volcanic aerosol optical depths following an eruption are typically less than 0.3 and decay with an e-folding time of about a year⁶. This figure shows that the aerosol can weakly compensate increases in DUV-DNA that result from ozone reductions. For instance, an aerosol optical

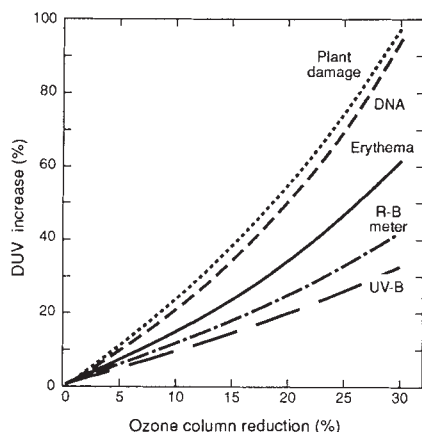


FIG. 1 Percentage increases in the daily dose of ultraviolet radiation (DUV) as a function of percentage reductions in the total ozone column. The curve labelled 'UV-B' gives the energy increase incident at the surface in the spectral range associated with UV-B radiation (290–320 nm). The curve labelled 'R-B meter' is produced by weighting the UV-B energy increases with the response function of a Robertson–Berger meter. The other curves are produced by weighting the UV-B energy increases with the appropriate biological action spectra. Ozone concentrations were depleted only in the region between 12 and 24 km.

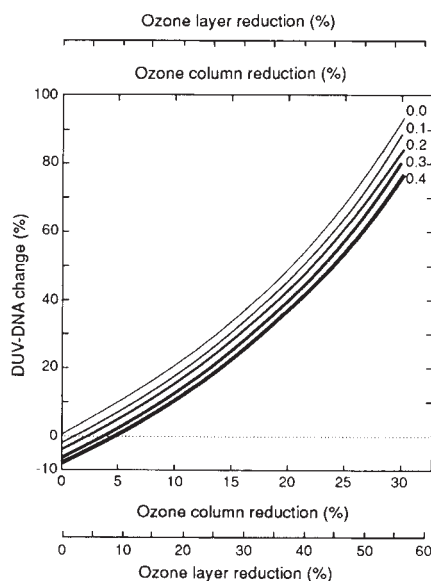


FIG. 2 Percentage changes in DUV-DNA (daily dose of DNA action spectra-weighted ultraviolet energy) are shown as a function of ozone depletions concurrent with volcanic aerosol loadings for the baseline case. Changes in DUV-DNA are shown for ozone column depletions ranging from 0 to 30% with concurrent changes in aerosol optical depths ranging from 0.0 to 0.4. Modelled ozone depletions and aerosol loadings were volumetrically uniform in the 12–24-km layer. The dual x-axes relate the percentage ozone depletion in the 12–24-km layer to the corresponding ozone column depletion in the baseline ozone profile; for example, a 15% layer depletion corresponds to an 8% change in the ozone column amount (as observed at some stations after the El Chichón eruption⁸). Any ozone column depletions as large as 30% would probably result from reductions extending outside the 12–24-km layer considered here.

depth of 0.2 can compensate completely only for small ozone column reductions of ~3%.

Soon after the March–April eruptions of El Chichón, observed optical depths were 0.3 or less, and a year later were less than 0.1 (ref. 6) when ozone column depletions of 8% were observed at northern latitudes. Estimates of the springtime ozone layer depletion caused by the El Chichón aerosol range from 2% to 10%. The increased DUV-DNA for a 2% layer depletion is effectively compensated by the increased aerosol amount, but the 10% layer depletion increases the DUV-DNA by ~8%. The optical depths immediately following the Mount Pinatubo eruption are slightly larger than those after El Chichón, about 0.4 or less²⁸; estimates of zonal-scale ozone layer depletions caused by the aerosol are 2–4%. For optical depths of 0.1–0.2, increases in DUV-DNA from these layer depletions are compensated for by the aerosol. But observations in Colorado indicated a short-term regional-scale ozone column depletion in November that was 20% lower than previous averages. Such an ozone depletion could increase the DUV-DNA by up to 45%.

These results suggest that major volcanic eruptions may increase DUV-BE. The net change of DUV-BE following such an eruption depends primarily on the magnitude of the ozone depletion, and only weakly on the amount of volcanic aerosol present. This conclusion agrees with that of another assessment³⁹. The magnitude of the ozone depletion depends on the ozone-destroying chemistry following an eruption, and on other factors (such as the quasi-biennial oscillation or El Niño/Southern Oscillation). For a more accurate determination of DUV-BE changes following a large eruption, we need a better understanding of the relation between the amount of aerosol present and the ozone depletion produced. If the activation of inert stratospheric chlorine is found to cause large ozone depletions after large eruptions, continued anthropogenic chlorine emissions will increase the possibility of even larger ozone depletions and DUV-BE increases following future eruptions. Even if anthropogenic trace-gas emissions are restricted in accordance

with the London protocol, ozone sensitivity to volcanic eruptions could be twice as large in 2010 as when El Chichón erupted³³.

The biological impact of such ozone depletions depends on the magnitude of the absolute change in DUV-BE. This change depends on the magnitude of the relative change (as given here) and on the background amount of DUV-BE at the time of the change. The ozone depletions tend to be largest in the winter–spring period when the climatological DUV-BE is low, thus lowering the risk of ultraviolet damage. Ozone depletions in the winter–spring period may still be important for organisms that are sensitive to exposure during that time. An organism may be sensitive if it is already stressed by current UV-B levels, as may be the case for phytoplankton^{34,35}; if it is in its earlier, more sensitive stages of development, as has been shown with anchovy larvae³⁶ and some soybean cultivars during the transition stage between vegetative and reproductive growth³⁷; or if it is stressed by other environmental or ecological factors. Greater UV-B stresses are likely in the future if the downward trends of stratospheric ozone continue without compensation, particularly in mid- and high latitudes³⁸.

Although the El Chichón and Mount Pinatubo eruptions are the largest of this century, they are small compared with other historically recent eruptions. For instance, the eruption of Tambora in 1815 injected an estimated 100–200 Mt of SO₂ into the stratosphere³, a mass 5–10 times larger than that injected by Pinatubo. If the theories showing that chlorine activation by volcanic aerosols can efficiently catalyse ozone destruction are proven correct, an eruption such as Tambora could have extremely deleterious consequences in our increasingly chlorine-enriched atmosphere. □

Received 11 March; accepted 2 July 1992.

- Hofmann, D. J. & Solomon, S. *J. geophys. Res.* **94**, 5029–5041 (1989).
- Turco, R. P. & Hamill, P. *Ber. Bunsenges. Phys. Chem.* **96**, 323–334 (1992).
- Pinto, J. P., Turco, R. P. & Toon, O. B. *J. geophys. Res.* **94**, 11165–11174 (1989).
- Prather, M. *J. geophys. Res.* **97**, 10187–10191 (1992).
- Michelangeli, D. V., Allen, M. & Yung, Y. L. *J. geophys. Res.* **94**, 18429–18444 (1989).
- Hofmann, D. *J. Rev. Geophys.* **25**, 743–759 (1987).
- Angell, J. K., Korshover, J. & Planet, W. G. *Mon. Weath. Rev.* **113**, 641–646 (1985).
- Bojkov, R. D. *Mon. Weath. Rev.* **115**, 2187–2201 (1987).
- Angell, J. K. *J. appl. Meteor.* **27**, 91–97 (1988).
- Chandra, S. & Stolarski, R. S. *Geophys. Res. Lett.* **18**, 2277–2280 (1991).
- Bluth, G. S. J., Doiron, S. D., Schnetzler, C. C., Krueger, A. J. & Walter, L. S. *Geophys. Res. Lett.* **19**, 151–154 (1992).
- Ackerman, T. P., Liou, K. N. & Leovy, C. B. *J. appl. Meteor.* **15**, 28–35 (1976).
- Cuzzi, J. N., Ackerman, T. P. & Helmle, L. C. *J. atmos. Sci.* **39**, 917–925 (1982).
- Goody, R. M. *Atmospheric Radiation I, Theoretical Basis* (Oxford Univ. Press, 1964).
- Molina, L. T. & Molina, M. J. *J. geophys. Res.* **91**, 14501–14508 (1986).
- World Meteorological Organization *Atmospheric Ozone 1985*, Vol. 1, Rep. 16 (Global Ozone Research and Monitoring Project, Geneva, 1986).
- Toon, O. B. & Pollack, J. B. *J. appl. Meteor.* **15**, 225–246 (1976).
- Shettle, E. P. & Fenn, R. W. *Air Force Geophys. Lab. Rep. AFGL-TR-79-0214*, NTIS accession no. ADA-085-951 (US National Technical Information Service, 1979).
- Lubin, D., Frederick, J. E., Booth, C. R., Lucas, T. & Neuschuler, D. *Geophys. Res. Lett.* **16**, 783–785 (1989).
- Lubin, D. & Frederick, J. E. *J. geophys. Res.* **95**, 13883–13889 (1990).
- Coillit, T. P. in *Impact of Global Climatic Changes on Photosynthesis and Plant Productivity* (Oxford, New Delhi, 1991).
- Setlow, R. B. *Proc. Natn. Acad. Sci. USA* **71**, 3363–3366 (1974).
- Caldwell, M. M., Camp, L. B., Warner, C. W. & Flint, S. D. in *Stratospheric Ozone Reduction, Solar Ultraviolet Radiation and Plant Life*, 87–111 (Springer, Berlin, 1986).
- Parrish, J. A., Jaenicke, K. F. & Anderson, R. R. *Photochem. Photobiol.* **36**, 187–191 (1982).
- Berger, D. S. *Photochem. Photobiol.* **24**, 587–593 (1976).
- Frederick, J. E. & Lubin, D. *Photochem. Photobiol.* **47**, 571–578 (1988).
- Stamnes, K., Slusser, J., Bowen, M., Booth, C. & Lucas, T. *Geophys. Res. Lett.* **17**, 2181–2184 (1990).
- Valero, F. P. J. & Pilewskie, P. *Geophys. Res. Lett.* **19**, 163–166 (1992).
- McCormick, M. P., Swisler, T. J., Fuller, W. H., Hunt, W. H. & Osborn, M. T. *Geophys. Res. Lett.* **23**, 187–221 (1994).
- US Standard Atmosphere (NOAA-S/T 76-1562) (US Government Printing Office, Washington, DC, 1976).
- Blumthaler, M. & Ambach, W. *Photochem. Photobiol.* **48**, 85–88 (1988).
- Michelangeli, D. V. *et al. J. geophys. Res.* **97**, 865–874 (1992).
- Braeser, G. P., Granier, C. & Walters, S. *Nature* **348**, 626–628 (1990).
- Worrest, R. C. in *The Role of Solar Ultraviolet Radiation in Marine Ecosystems*, 429–457 (Plenum, New York, 1982).
- Smith, R. C. *et al. Science* **255**, 952–959 (1992).
- Hunter, J. R., Kaupp, S. E. & Taylor, J. H. in *The Role of Solar Ultraviolet Radiation in Marine Ecosystems*, 459–497 (Plenum, New York, 1982).
- Teramura, A. H. & Sullivan, J. H. *Am. J. Bot.* **74**, 975–979 (1987).
- Madronich, S. *Geophys. Res. Lett.* **19**, 37–40 (1992).
- United Nations Environment Programme *Environmental Effects of Ozone Depletion: 1991 update* (United Nations Environment Programme, Nairobi, Kenya, 1991).

ACKNOWLEDGEMENTS. We thank R. M. Klein and W. H. Brune for discussions, and D. Greeman for comments on the manuscript. Portions of this research were funded through NASA and NSF. A.V. acknowledges support from the NASA Global Change Graduate Fellowship Program.

Volcanic winter and accelerated glaciation following the Toba super-eruption

Michael R. Rampino*† & Stephen Self‡

* Earth Systems Group, Applied Science Department, New York University, New York, New York 10003, USA

† NASA, Goddard Institute for Space Studies, New York, New York 10025, USA

‡ Department of Geology and Geophysics, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA

THE eruption of Toba in Sumatra 73,500 years ago was the largest known explosive volcanic event in the late Quaternary¹. It could have lofted about 10^{15} g each of fine ash and sulphur gases to heights of 27–37 km, creating dense stratospheric dust and aerosol clouds. Here we present model calculations that investigate the possible climatic effects of the volcanic cloud. The increase in atmospheric opacity might have produced a ‘volcanic winter’²—a brief, pronounced regional and perhaps hemispheric cooling caused by the volcanic dust—followed by a few years with maximum estimated annual hemispheric surface-temperature decreases of 3–5 °C. The eruption occurred during the stage 5a–4 transition of the oxygen isotope record, a time of rapid ice growth and falling sea level³. We suggest that the Toba eruption may have greatly accelerated the shift to glacial conditions that was already underway, by inducing perennial snow cover and increased sea-ice extent at sensitive northern latitudes. As the onset of climate change may have helped to trigger the eruption itself⁴, we propose that the Toba event may exemplify a more general climate–volcano feedback mechanism.

The Toba pyroclastic deposits on Sumatra near to the source (equal to 2,000 km³ of magma) have been radiometrically dated at 73.5 ± 3.5 kyr and 73 ± 4 kyr before present (ref. 1). The widespread Toba ash layer (800 km³ of magma)¹ occurs in Indian Ocean deep-sea cores within the oxygen isotope stage 5a–4 transition³. The duration of eruption has been estimated as up to two weeks⁵, and the eruptive flux as $\sim 7 \times 10^9$ kg s⁻¹, implying eruption cloud heights from 27 to 37 km (ref. 6).

Petrological studies suggest that $\sim 3 \times 10^{15}$ g of H₂S and SO₂ (convertible to $\sim 1 \times 10^{16}$ g of H₂SO₄ aerosols) could have been

released from the erupted magma¹. Although the intrinsic sulphur content of silicic magmas is low², efficient degassing and the large volume of magma erupted combine to give an enormous volatile release. Extrapolation from smaller historical eruptions² suggests a lower amount of aerosol (1.5×10^{15} g) for Toba, along with $\sim 2 \times 10^{16}$ g of fine (<2 µm) dust. Condensation and coagulation in aerosol clouds may considerably reduce the amount of long-lived H₂SO₄ aerosols⁷. At present, however, there are no data on the behaviour of H₂SO₄ aerosols in such dense clouds.

The amount of stratospheric water available to convert SO₂ to H₂SO₄ may limit aerosol loading. At present, $\sim 4 \times 10^{15}$ g of ambient water⁸ might be available in the form of H₂O and CH₄. Toba may have released as much as 5.4×10^{17} g of H₂O (ref. 1), so enough could have been present in the stratosphere to convert the sulphur gases emitted by Toba into $\sim 1 \times 10^{16}$ g of H₂SO₄ aerosols. To be conservative, however, and in agreement with extrapolations from historic eruptions, we assume that only 10% of this amount of aerosols actually formed in the Toba cloud, or $\sim 1 \times 10^{15}$ g of H₂SO₄ aerosols.

The 1815 Tambora eruption (with an estimated 2×10^{14} g of stratospheric aerosols) may be related to an observed surface temperature decrease of ~ 0.7 °C in the Northern Hemisphere in the year following the eruption². A linear relationship between hemispheric cooling and aerosol loading would predict cooling by ~ 3.5 °C in the year following the eruption from an aerosol loading of 1×10^{15} g. Other empirical methods⁹ give estimates in the range of 3 to 5 °C.

A stratospheric aerosol burden for Toba of 1×10^{15} g yields a global aerosol optical depth (τ) of about 10 (ref. 2). (Optical depth $\tau = -\ln(I/I_0)$; where I_0 and I are the initial and final light intensity, respectively. For volcanic aerosols and dust, τ is largely dependent on the single scattering albedo (ω) of the cloud, whereas for soot (smoke) clouds, the absorption optical depth τ_a is primarily a function of the high absorptivity of the soot particles in the visible (fluffy soot also absorbs in the infrared).) A stratospheric injection of 10% of the estimated fine dust generated would produce a dust-cloud optical depth of ~ 10 for a few months. The low-latitude position of Toba would probably have led to efficient injection of dust and volatiles into the stratosphere in both hemispheres². The estimated optical depth increase from the dust and aerosols is equivalent in visible opacity to smoke-cloud absorption optical depths of ~ 2 to 3, used in nuclear-winter scenarios for a full-scale nuclear war^{10–12}.

In simulations of the first 1 to 3 months of a nuclear winter (for a worst-case July smoke injection) global climate models

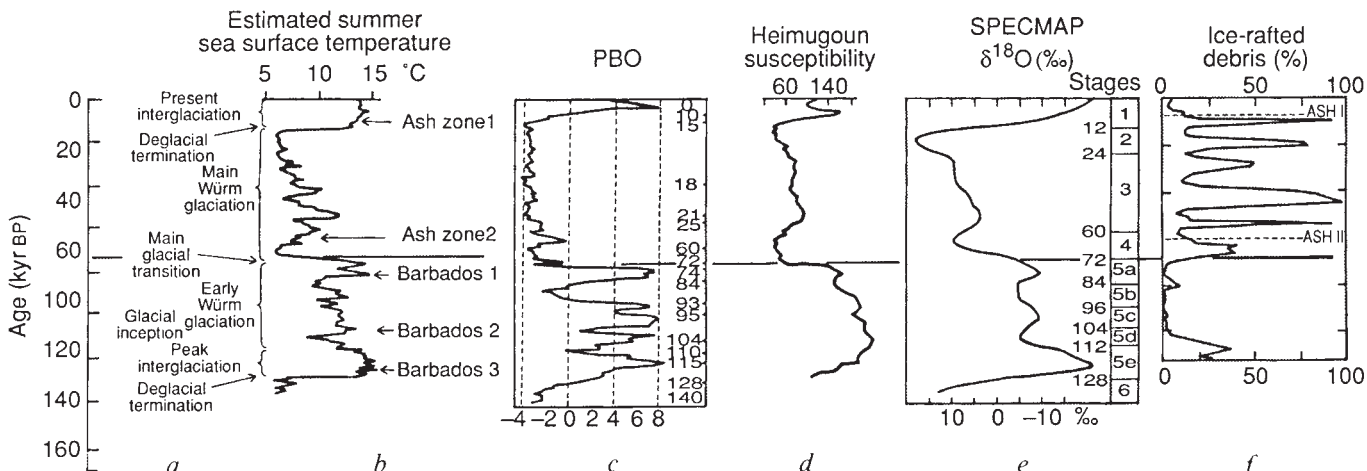
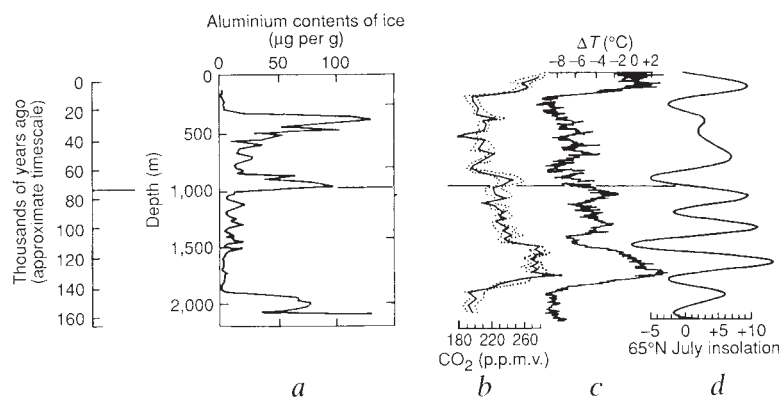


FIG. 1 Climate indicators for the past 140 kyr. *a*, Climatic states³³. *b*, Reconstructed summer SST (planktonic foraminiferal assemblage data), deep-sea core V23-82, North Atlantic¹⁵. Dates: Ash Zone 1 (9.5 kyr), Ash Zone 2 (65 kyr), Barbados 1 Terrace (~85 kyr), Barbados 2 (~110 kyr) and Barbados 3 (~125 kyr). *c*, Palaeobioclimatic operator (PBO), a measure of the ‘best possible’ climate profile based on multivariate statistical analysis of Les Echets, France pollen records. Higher PBO indicates warmer climate.

The vertical scale is nonlinear⁴⁵. *d*, Magnetic susceptibility, Heimugou loess section, Central China. Loess in palaeosols formed in warm intervals has higher susceptibility than loess deposited in cold intervals⁴⁶. *e*, SPECMAP $\delta^{18}\text{O}$ scale. Stages and dates of boundaries shown at right³. $\delta^{18}\text{O} = [(^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{standard}}] - 1$, where standard is PDB (cretaceous belemnite carbonate). *f*, Percentage of ice-rafted detritus in North Atlantic deep-sea core Me69-17. Depth in metres¹⁴.

FIG. 2 Vostok ice-core climate records²¹, and climatic forcings for the past 160 kyr. For ice-core data, timescale is approximate. *a*, Aluminium content of ice, indicating continental wind-blown dust. Vertical scale is in metres depth. *b*, Atmospheric CO₂ from trapped air bubbles. *c*, Temperature deviations (°C) from present, based on δD measurements. *d*, July insolation at 65°N (per cent deviation from present)⁴⁸.



(GCMs) predict that land temperatures in the latitude zone 30° to 70° N could range from ~5 to ~15 °C colder than normal, with widespread immediate hard freezes in mid-latitudes, and possibly considerable precipitation decreases¹⁰. At lower latitudes, GCM simulations suggest immediate coolings of ~10 °C (ref. 11). During the first 1 to 3 years, global average τ_a of 0.5 could persist, with cooling of 5 °C in areas covered by the smoke cloud. Ocean-surface cooling of ~2 to 6 °C may extend for several years¹⁰. The persistence of significant soot optical depths for 1–3 years might lead to longer-term (decadal) cooling, primarily through increased snow cover and sea ice, and perturbed sea surface temperatures^{10–11}.

The injection of massive amounts of volcanic dust into the stratosphere by a super-eruption such as Toba, with aerosol optical depths of ~10, might be expected to cause similar immediate surface cooling, creating a 'volcanic winter'². Volcanic dust has a shorter residence time in the atmosphere (3–6 months) than soot^{10,11}, and hence the effects might be more limited in area and in time. But the coincident injection of ~10¹⁵ g of sulphur volatiles into the stratosphere, and the time required for the formation and spread of the resulting H₂SO₄ aerosols, should extend the period of increased atmospheric opacity and surface cooling. A maximum e-folding residence time for the aerosols in the stratosphere of about a year² (ignoring possibly important self-limiting mechanisms) would lead to several years with optical depths ≥ 0.1 .

The Toba eruption occurred during a major Late Quaternary climate event, the $\delta^{18}O$ stage 5a-4 boundary, which is the interval of most rapid ice accumulation, and the main Northern Hemisphere warm-cold transition of the last glacial cycle^{3,13} (Figs 1 and 2). The transition is marked by a drop in sea level of ~40 m in less than 7,000 years, the first significant peak in ice-raftered detritus in the North Atlantic during the last glacial cycle¹⁴, and a decrease in seasonal sea surface temperatures in the North Atlantic¹⁵ by 10 °C in <5,000 years. Other climate indicators also suggest that the glacial cooling began very abruptly^{16–18} at ~74 kyr; changes in ocean bottom-water circulation in the Atlantic may have involved a shut-down of North Atlantic Deep Water formation^{19,20}, briefly in operation during substage 5a.

At ~74 kyr, Milankovitch parameters were such as to favour the growth of Northern Hemisphere ice sheets, with low summer insolation at high northern latitudes (Fig. 2). Average atmospheric p_{CO_2} decreased from ~245 p.p.m. to ~195 p.p.m. mostly during the final phase of the cooling ~71 to 62 kyr (ref. 21). Atmospheric methane shows a decrease from about 600 to 400 parts per 10⁹ (p.p.b.) which is only a minor climate forcing²². Increased atmospheric dust content during glacial stages²³ has been suggested to cause global cooling of up to 2–3 °C: the Vostok ice core shows a dust peak at ~71 kyr (Fig. 2). An increase in non-sea-salt sulphur in the Vostok core²⁴ at the stage 5a-4 boundary suggests a possible minor cooling from increased cloud cover of ~0.6 °C.

Much of the large increase in ice volume occurred next to a

warm subpolar ocean¹⁵. By stage 4, ice volume reached more than 60% of the full glacial maximum of stage 2, with snow and ice fields southward in eastern Canada to ~50 °N (ref. 16). These conditions could be interpreted as supporting the concept of rapid 'glacierization'^{25,26}, in which summer cooling leads to permanent snowfields and resulting snow-albedo feedback.

A period during which mean daily summer temperatures were lower than present over the central plateau of Quebec and Labrador, perhaps accompanied by increased snowfall, might favour the formation of perennial snow cover^{25–28}. Tree-ring data from Northern Quebec show mean summer temperatures for the past 250 years of 9.5 °C, with a lowering of 3.5 °C (to ~6 °C) in 1816 after the Tambora eruption²⁹. Northern Hemisphere temperatures decreased by ~0.7 °C in the year after Tambora², so that the summer temperature decline in Quebec in 1816 shows roughly a fivefold amplification of hemispheric cooling (although this extreme cooling could be partly a dynamic effect of regional blocking and meridional circulation³⁰). If these same factors were important at 74 kyr, then a 3 °C Northern Hemisphere cooling after Toba might have led to Quebec summer temperatures 10–15 °C colder than at present for 2 or 3 years.

Lowering of the freezing line should have a direct effect on annual snow accumulation. July temperatures on the east shore of Hudson Bay at present average 14.2 °C, and the freezing altitude (0° isotherm) averages ~2,500 m (ref. 31). An aerosol cloud like that of Toba might cause the atmospheric freezing altitude to descend to ground level under present July climate conditions, an even more likely occurrence under conditions of increased snow and ice cover and sea-ice extent.

According to the Goddard Institute for Space Studies (GISS) GCM, volcanic aerosols that caused an optical-depth perturbation of 0.1 for 5 years led to a global surface cooling of ~0.5 °C, with a reduction in summer temperatures in Canada at ~55° N of ~2 °C, or 4 times the global average (D. Rind, personal communication). If high-latitude seasonal amplification is roughly linear with temperature decrease, then a 3 °C global cooling after the Toba eruption could have led to a 12 °C reduction in summer temperatures for perhaps ~2 to 3 years in the areas of growth of the Laurentide ice sheet, in good agreement with empirical estimates.

Results of two-dimensional climate models coupled with ice-sheet models^{32,33} are in agreement with the inferred isotopic and sea-level changes at the 5a-4 boundary³⁴, but these models are highly parameterized, and do not rule out a contribution from volcano-induced cooling. By contrast, GCM experiments give mixed results^{27,35,36}. Some GCM runs using versions of the NCAR Community Climate Model suggest that imposition of snow cover in the Northern Hemisphere under present conditions can lead to retention and increase in snow depth in glaciation-sensitive areas³⁵. Experiments using the GISS GCM³⁶, however, were unable to initiate the growth of Northern Hemisphere ice sheets even with summer insolation fields lower than that of 74 kyr. The effects of considerable pre-existing land ice

and sea ice on further ice accumulation are not well determined, and the mixed GCM results suggest differences in model sensitivity to factors involved in ice accumulation. Until these are better understood, no firm conclusions should be drawn about ice-growth conditions.

The detailed record of climate and $\delta^{18}\text{O}$ during the stage 5a-4 transition reveals that although sea-level lowering³ began before the Toba eruption, North Atlantic surface-ocean temperatures remained warm³⁷, and ice-sheet growth was beginning to slow³. It may be significant, therefore, that the Toba eruption apparently coincided with a precipitous decrease in North Atlantic surface temperatures and global sea level^{3,34}. Increased sea ice and snow cover following the eruption may have provided the extra 'kick' that caused the climate system to switch from warm to cold states^{38,39}. Other very large Quaternary eruptions during warm-cold transitions show a similar timing^{40,41}, and there is also some evidence of increased volcanism during cold-warm transitions⁴¹. The connection may come through the effects of water loading and unloading on magma chambers during sea-level fluctuations⁴, which may explain the apparent relationship between volcanic activity and Milankovitch parameters: Late Quaternary ash layers in deep-sea cores from the Mediterranean⁴² and the Indian Ocean⁴³ have a periodicity of ~23 kyr similar to the Earth's precessional period, and Icelandic⁴⁴ and Japanese eruptions (K. Caldeira, personal communication) apparently follow a 100-kyr cycle.

Although correlations have often been claimed between Quaternary climate change and volcanism⁴, it was thought that because of the short-lived nature of volcanic aerosols, and the control of the timing of climatic change by Milankovitch cycles, such correlations were unrelated coincidence. Our results may provide a physical explanation: if climatic change can trigger eruptions⁴, then short-term volcanic winters caused by dust and aerosols from large eruptions such as Toba might provide positive feedback for cooling at critical times during warm-cold transitions. Volcanic aerosols may also constitute a negative feedback during glacial terminations, contributing perhaps to brief episodes of cooling and glacial readvance such as the Younger Dryas Interval⁴¹. □

Received 10 December 1991; accepted 9 July 1991.

1. Chesner, C. A. *et al. Geology* **19**, 200-203 (1991).
2. Rampino, M. R., Self, S. & Stothers, R. B. *Rev. Earth planet. Sci.* **16**, 73-99 (1988).
3. Martinson, D. G. *et al. Quat. Res.* **7**, 1-29 (1987).
4. Rampino, M. R., Self, S. & Fairbridge, R. W. *Science* **206**, 826-828 (1979).
5. Ledbetter, M. T. & Sparks, R. S. J. *Geology* **7**, 240-244 (1979).
6. Woods, A. W. & Wohletz, K. H. *Nature* **350**, 225-227 (1991).
7. Pinto, J. P., Turco, R. P. & Toon, O. B. *J. geophys. Res.* **94**, 11165-11174 (1989).
8. Stothers, R. B., Wolff, J. A., Self, S. & Rampino, M. R. *Geophys. Res. Lett.* **13**, 725-728 (1986).
9. Palais, J. M. & Sigurdsson, H. *Am. geophys. Un. geophys. Monogr.* **52**, 31-52 (1989).
10. Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B. & Sagan, C. *Science* **247**, 166-170 (1990).
11. Covey, C., Schneider, S. H. & Thompson, S. *Nature* **308**, 21-25 (1984).
12. National Research Council *The Effects on the Atmosphere of a Major Nuclear Exchange* (National Academy of Sciences, Washington DC, 1985).
13. Ruddiman, W. F. & McIntyre, A. *Science* **204**, 173-175 (1979).
14. Heinrich, H. *Quat. Res.* **29**, 142-152 (1988).
15. Sancetta, C., Imbrie, J., Kipp, N. G., McIntyre, A. & Ruddiman, W. F. *Quat. Res.* **2**, 363-367 (1972).
16. Koerner, R. M. & Fisher, D. A. in *Late Quaternary Environments: Eastern Canadian Arctic, Baffin Bay and West Greenland* (ed. Andrews, J. T.) 309-327 (Allen and Unwin, London, 1985).
17. Woillard, G. & Mook, W. G. *Science* **215**, 159-161 (1982).
18. Guiot, J. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **80**, 49-69 (1990).
19. Streeter, S. S. & Shackleton, N. J. *Science* **203**, 168-171 (1979).
20. Jouzel, J. *et al. Nature* **329**, 403-407 (1987).
21. Barnola, J. M. *et al. Nature* **329**, 408-414 (1987).
22. DeAngelis, M., Barkov, N. I. & Petrov, V. N. *Nature* **325**, 318-321 (1987).
23. Harvey, L. D. *Nature* **334**, 333-335 (1988).
24. Legrand, M. R., Delmas, R. J. & Charlson, R. J. *Nature* **334**, 418-420 (1988).
25. Loewe, F. *Arct. Alp. Res.* **3**, 331-344 (1971).
26. Andrews, J. T. & Barry, R. G. *Rev. Earth planet. Sci.* **6**, 205-228 (1978).
27. Williams, J. *J. appl. Meteorol.* **14**, 137-152 (1975).
28. Koerner, R. M. *Quat. Res.* **13**, 153-159 (1980).
29. Jacoby, G. C., Ivanciu, I. S. & Ulan, L. D. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **64**, 69-78 (1988).
30. Schneider, S. H. *Clim. Change* **5**, 111-113 (1983).
31. Wilson, C. *Syllogeus* **55**, 147-190 (1985).
32. Fichet, T. *et al. Phil. Trans. R. Soc. Lond. A* **339**, 249-261 (1989).
33. Deblonde, G. & Peltier, W. R. *Clim. Dynam.* **5**, 103-110 (1990).
34. Chappell, J. & Shackleton, N. J. *Nature* **324**, 137-140 (1986).
35. Oglesby, R. J. *Clim. Dynam.* **4**, 219-235 (1990).
36. Rind, D., Peteet, D. & Kukla, G. *J. geophys. Res.* **94**, 12851-12871 (1989).
37. Ruddiman, W. F., McIntyre, A., Niebler-Hunt, V. & Durazzi, J. T. *Quat. Res.* **13**, 33-64 (1980).
38. Matteucci, G. *Clim. Dynam.* **6**, 67-81 (1991).

39. Miller, G. H. & Kaufman, D. S. *Norsk Geol. Tidsskr.* **71**, 149-151 (1991).
40. Ruddiman, W. F. & McIntyre, A. *Geol. Soc. Am. Bull.* **93**, 1273-1279 (1982).
41. Dawson, A. *Ice Age Earth: Late Quaternary Geology and Climate*, 180-198 (Routledge, London, 1991).
42. Paterne, M. *et al. Earth planet. Sci. Lett.* **98**, 166-174 (1990).
43. Paterne, M. *et al. IAVCEI Progr. Abstr., Gen. Assembly, Vienna, IUGG* **20**, 6 (1991).
44. Sioholm, J., Sejrup, H. P. & Furnes, H. *J. Quat. Sci.* **6**, 159-173 (1991).
45. Guiot, J. *et al. Nature* **338**, 309-313 (1989).
46. Heller, F., Xiuming, L., Tungsheng, L. & Tongchun, X. *Earth planet. Sci. Lett.* **103**, 301-310 (1991).
47. Ninkovich, D. *et al. Nature* **276**, 574-577 (1978).
48. Berger, A. *Quat. Res.* **9**, 139-167 (1978).

ACKNOWLEDGEMENTS. We thank K. Caldeira, C. Fletcher, D. Peteet, W. F. Ruddiman, S. H. Schneider, R. B. Stothers and G. P. L. Walker for reviews, and D. Rind, W. I. Rose, H. Sigurdsson and A. W. Woods for discussions. This work was supported in part by grants from the U.S. DOE, National Institutes for Global Environmental Change, NASA, the Center for Global Habitability at Columbia University (M.R.) and NATO (S.S.).

Evidence for melt segregation towards fractures in the Horoman mantle peridotite complex

Natsuko Takahashi

Geological Institute, Faculty of Science, University of Tokyo, Tokyo 113, Japan

MELT segregation from a solid matrix in the upper mantle is the first step in the formation of magmas. The mechanisms of melt segregation proposed so far have been based on two different physical models: the percolation of interstitial melt driven by a density difference between melt and matrix, associated with compaction and deformation of the matrix¹; or the suction of interstitial melt into fractures from a surrounding porous matrix²⁻⁴. Although the latter process was originally invoked to explain the occurrence of dunite and pyroxenite-gabbro dykes with zones depleted in melt component on both sides^{2,5-7}, most of these zones are discordant, small in scale and clearly formed by reaction between peridotite and basaltic melt⁸. I have suggested⁹ that thick, concordant dunite zones in the most depleted peridotite layers of the Horoman peridotite complex formed by suction of partial melt towards fractures later filled by dunite. Here I report a detailed mineralogical variation across the dunite which, when set in its geological and petrographic context, provides clearer evidence for melt segregation by dynamic forcing¹, rather than passive percolation. This mechanism of melt segregation may play an important role in the formation of primary magmas with low production rate and large geochemical variability.

The Horoman peridotite complex⁹⁻¹⁴ in Hokkaido, northern Japan, lies at the southern end of the Hidaka metamorphic belt¹⁵. The complex is stratified and has a total thickness of ~3,000 m (refs 11, 12). It is composed of the Main Harzburgite-Lherzolite suite (MHL), with a subordinate spinel-rich dunite-wehrlite suite (SDW) and a minor banded dunite-harzburgite suite (Fig. 1)^{9,14}.

The MHL is composed of harzburgite and lherzolite, which are grouped into harzburgite, symplectite-free spinel lherzolite, symplectite-bearing spinel lherzolite and plagioclase lherzolite on the basis of mineral content and presence or absence of symplectites consisting of intergrowths of ortho- and clinopyroxene plus spinel ('two-pyroxene/spinel' symplectites). These four rock types show gradational changes in modal composition and bulk and mineral chemistries: the modal content of clinopyroxene and orthopyroxene, and the bulk content of CaO and Al₂O₃ decrease, whereas the forsterite content of olivine (Fo) and the Cr/(Cr+Al) atomic ratio (Cr#) of spinel increase from plagioclase lherzolite to harzburgite through symplectite-bearing and symplectite-free spinel lherzolite^{9,14}. The variations are cyclic in the field, giving rise to the stratified structure of the Horoman complex. The wavelength of each cycle ranges from several meters to a few hundred metres. The

SDW consists mainly of spinel-rich (up to 5 modal%) dunite with small amounts of wehrlite, and always occurs in harzburgite layers.

Two features of the Horoman complex constrain melting and melt-segregation processes in the upper mantle. One is the presence of symplectites in the fertile peridotites, and the other is the occurrence of SDW dunite layers in harzburgite zones.

Two-pyroxene/spinel symplectite is most frequently observed in symplectite-bearing spinel lherzolite. It also occurs in orthopyroxene porphyroclasts of plagioclase lherzolites. The textural and bulk chemical composition of the symplectite indicate that it is a garnet pseudomorph^{9,16}. Plagioclase in the plagioclase lherzolite is inferred to be a product of subsolidus reaction between the pyroxenes and spinel, mainly in the symplectite, because plagioclase is restricted to elongated fine-grained aggregates^{9,16}. These facts suggest that the plagioclase lherzolite was only slightly molten, if at all. The systematic correlation between textural and chemical variations of symplectite and degree of refractoriness indicates that melting and melt segregation in the Horoman peridotites took place almost contemporaneously with the decompression reaction between garnet and olivine⁹. If the source garnet lherzolite had a uniform composition, then localized melting and melt segregation must have occurred to produce residual peridotites with variable depletion in melt components. The mechanism of the melt-segregation process, which may be strongly coupled with melting, can be constrained from the occurrence of dunite (SDW) in the harzburgite layers of MHL and mineralogical variation in the surrounding peridotites.

The SDW dunite occurs as concordant layers with sharp boundaries and is always sandwiched between harzburgite layers

of MHL^{9,14} (Fig. 1). The thickness of the dunite layers ranges from 20 cm to 15 m. The modal abundance of clinopyroxene and orthopyroxene in the surrounding harzburgite decreases towards the SDW dunite layer. At the contact, harzburgite becomes almost clinopyroxene-free, and orthopyroxene suddenly disappears in the harzburgite, and abundant spinel grains appear in the dunite. The thickest SDW layer is ~15 m thick and has a wehrlitic centre (Fig. 2). The boundaries with surrounding harzburgite are sharp not only in modal abundance but also in mineral chemistry and marked by an abrupt change in the forsterite (Fo) and NiO content of olivine and in the chromium number Cr# of spinel (Fig. 2). By contrast, in the dunite layer the Fo and NiO contents of olivine decrease slightly towards the centre of the layer. These observations are common to all investigated SDW dunite layers and the surrounding MHL peridotites.

The dunite layer in depleted harzburgite is not a residue left after a high degree of melting, because it is more enriched in spinel and clinopyroxene than the harzburgite, and because the Fo content and NiO content of olivine and Cr# of spinel are lower than those of the depleted harzburgite. The SDW and surrounding depletion zone may not be the result of reaction

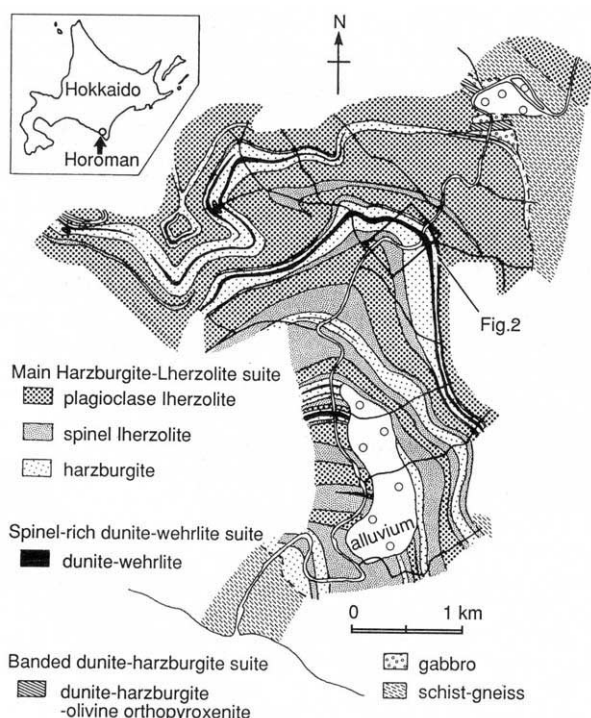


FIG. 1 Lithological map of the southern part of the Horoman peridotite complex, which is $\sim 8 \times 10$ km in size. The layers dip gently to the north in this area¹². The lithological variation in MHL is gradational. The boundary between harzburgite and lherzolite is at a volume ratio of clinopyroxene/(clinopyroxene + orthopyroxene) = 0.1 (ref. 25). The plagioclase lherzolite was identified as peridotite with fine-grained seam in the field. Zones of gabbro-peridotite bandings are observed in some upper horizons and very rarely in lower horizons^{11,12}. They are not shown because gabbro layers are much thinner than peridotites (usually tens of centimetres in thickness).

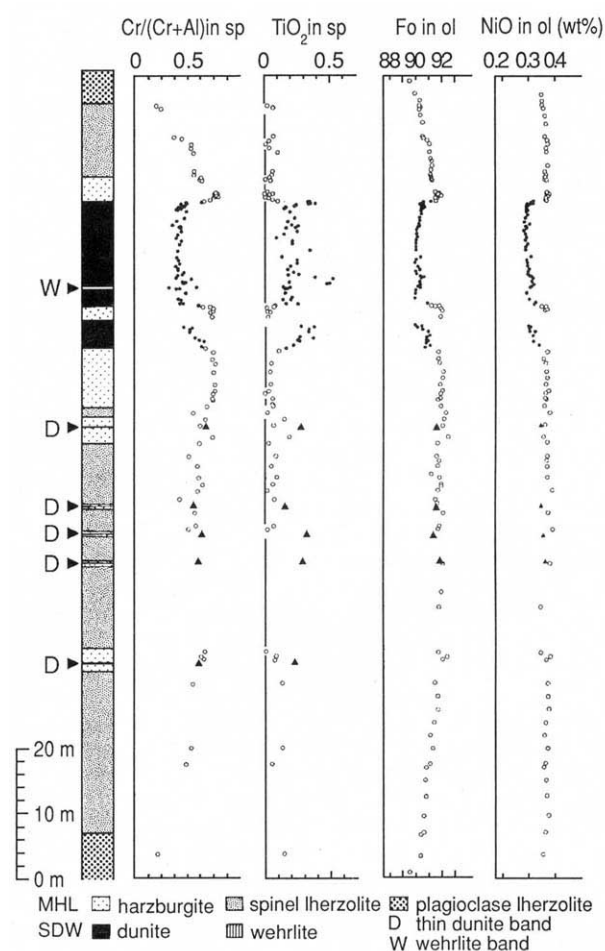
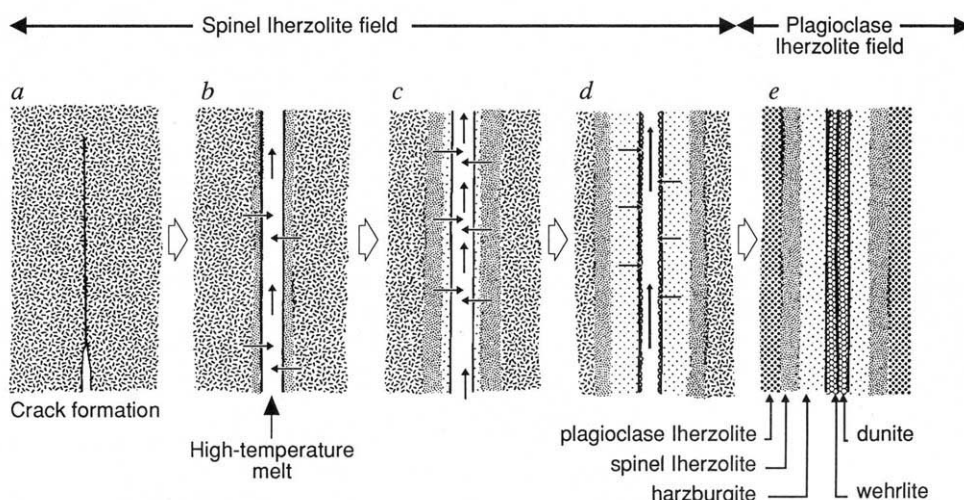


FIG. 2 Variation in mineral chemistry of the thickest dunite of SDW (●) and the surrounding MHL peridotites (○). The dunite layer can be traced for more than 4 km along strike in harzburgite layer (Fig. 1). ▲, Thin dunite layers associated with harzburgite. The NiO content of olivine (ol) was determined with a JEOL wavelength dispersive microprobe (733 MKII, Geological Institute, University of Tokyo) with 25 kV accelerating voltage and 100 s counting time. The Cr# of spinel (sp) and Fo content of olivine. In the MHL peridotite gradually increase towards the contact with the dunite layer on the both sides and abruptly decrease at the contact. In the uppermost 18 m, the mineralogical variations are much larger than the range of all data from the abyssal peridotites²⁶.

FIG. 3 One-dimensional melt-segregation model governed by suction associated with local melting along a fracture during continuous ascent of mantle peridotite. The crack is assumed to be vertically oriented. Arrows indicate inferred direction of migrating intergranular melt. *a*, Crack formation in the fertile lherzolite in spinel peridotite stability field. *b*, High-temperature melt of deeper origin is transported along this crack, causing local melting of the host peridotite near the crack. Because of melting, two-pyroxene-spinel symplectite begins to disappear near the crack, but is still preserved away from the crack, where peridotites have little or no partial melt. The partial melt begins to segregate towards the crack driven by the pressure difference between the crack and the host peridotite. *c*, Melt segregation towards the crack causes selective depletion of the surrounding peridotite. *d*, During subsequent cooling, olivine crystallizes from the melt forming dunite on the crack wall. *e*, Solidification is further promoted on cooling, and the wehrlitic centre



crystallizes last. On entering the plagioclase lherzolite facies, reaction between pyroxenes and spinel produces a fine-grained aggregate of olivine-plagioclase in fertile peridotite away from the dunite.

between melt and peridotite as in the case of the Trinity complex, where the replacive origin of dunite is well documented^{5,6,8}. The replacive dunites and surrounding peridotites in the Trinity complex have the following differences from the Horoman complex. First, the modal abundance of orthopyroxene does not decrease systematically from plagioclase lherzolite to harzburgite in Trinity⁶, whereas in Horoman it continuously decreases towards harzburgite in contact with dunite from the most fertile plagioclase lherzolite, suggesting that a melt component has been subtracted. Second, in the Trinity complex the Fo content and Cr# of spinel monotonously increase or decrease towards dunite from the most fertile plagioclase lherzolite without any abrupt change. By contrast, in Horoman the Fo content of olivine and Cr# of spinel always increase towards the contact with dunite from plagioclase lherzolite, then suddenly decrease in dunite on crossing the boundary. These differences suggest that the surrounding depletion zone and dunite in the Horoman complex were not of replacive origin.

Olivine compositions of SDW plot along an equilibrium or fractional crystallization trend starting from a value similar to the surrounding harzburgite of the MHL. This fact, and the high Fe³⁺ and high TiO₂ contents of spinel, suggest a cumulus origin for the SDW dunite. The mode of occurrence of the SDW layers and the positive correlations between the thickness of the dunite layers and surrounding depletion zones, and between average Fo values of olivine in contacting dunite and harzburgite suggest that the SDW is a cumulate from the magma segregated from the surrounding MHL peridotite⁹. The planar shape of dunite layers, the fact that their thickness varies by up to three orders of magnitude, their cumulus origin, and the nature of the boundaries between the dunite layer and residual peridotites suggest that the dunites filled planar melt channels, probably fractures, in the upper mantle⁹.

The sharp contact between dunite and harzburgite corresponds almost exactly to the boundary at which an abrupt change in NiO and Fo contents in olivine takes place. This fact and the progressive enrichment of compatible elements in the surrounding harzburgite towards the dunite indicate that there was a significant component of melt flow towards a melt channel now filled by dunite. The boundary would be diffuse if melt migrated parallel to the contact, and the chemical discontinuities of elements with different compatibility would not coincide if melt flowed into harzburgite^{17,18}. This further suggests that there was a negative pressure gradient towards the melt channel³.

It has been proposed that crack formation and a local pressure

gradient induce melt extraction toward the crack⁴. One-dimensional modelling of this process suggests that a depletion zone may develop for the scale of a compaction length ranging from 100 to ~1,000 m (refs 1, 3). This range is consistent with the observed thickness of the depletion zones (tens to a few hundred metres).

Possible explanations for the local increase of degree of partial melting around the melt channels are lowering of the peridotite solidus by addition of fluid, heating by high-temperature melt flowing from deeper levels and pressure release as the cracks formed. All three mechanisms could operate simultaneously.

The introduction of exotic melt or fluid with a peculiar chemical signature could have caused geochemical modification of the surrounding peridotite, especially affecting concentrations of highly incompatible elements. Takazawa *et al.*¹⁹⁻²¹ have suggested that this occurs when melt percolates through the layers after their formation. But the behaviour of major and compatible elements (including heavy rare-earth elements) in the depletion zone, which are not much affected by chromatographic fractionation²⁰, indicates that most of the interstitial melt was originally extracted from the peridotite to the dunite zone, rather than being an exotic contribution.

Figure 3 illustrates a simplified one-dimensional model for the formation of a dunite layer in the Horoman complex. First, primary fertile lherzolite started to ascend from the garnet peridotite field into the spinel peridotite field, associated with reaction between olivine and garnet. High-temperature melt from greater depths passed through the Horoman primary peridotite, probably following fractures. Its introduction resulted in local melting of the host peridotite near the fracture. Melt segregation towards the crack, driven by a pressure difference between melt in the crack and in the host peridotite, caused systematic depletion of the surrounding peridotite as a function of distance from the crack.

These processes could have taken place during the ascent of the Horoman complex along a pressure-temperature trajectory close to the peridotite solidus in the spinel peridotite stability field. One possible tectonic situation is where a mantle plume rose and spread horizontally^{22,23}. In this situation, vertical or steeply inclined cracks could have been formed, which were later transposed into gently inclined layering²⁴. The occurrence of cumulate dunite making an abrupt contact with the most depleted peridotite indicates that suction of partial melt by dynamic forcing is a possible mechanism for melt segregation in upwelling mantle material. □

Received 13 January; accepted 22 July 1992.

- McKenzie, D. J. *J. Petrol.* **25**, 713–765 (1984).
- Nicolas, A. & Jackson, M. *J. Petrol.* **23**, 568–582 (1982).
- Ribe, N. M. *Geophys. Res. Lett.* **13**, 1462–1465 (1986).
- Sleep, N. H. *J. geophys. Res.* **93**, 10255–10272 (1988).
- Quick, J. E. *J. geophys. Res.* **86**, 11837–11863 (1981).
- Quick, J. E. *Contrib. Miner. Petrol.* **78**, 413–422 (1981).
- Nicolas, A. *Structures of Ophiolites and Dynamics of Oceanic Lithosphere*. (Kluwer, Dordrecht, 1989).
- Kelemen, P. B. *J. Petrol.* **31**, 51–98 (1990).
- Takahashi, N. *J. Miner. Petr. Econ. Geol.* **86**, 199–215 (1991).
- Nagasaki, H. *J. Fac. Sci. Univ. Tokyo* **16**, 313–346 (1966).
- Niida, K. *J. geol. Soc. Jap.* **80**, 31–44 (1974).
- Niida, K. *J. Fac. Sci. Hokkaido Univ. Ser. IV* **21**, 61–81 (1984).
- Obata, M. & Nagahara, N. *J. geophys. Res.* **92**, 3467–3474 (1987).
- Takahashi, N. in *Ophiolite Genesis and Evolution of Oceanic Lithosphere* (eds Peters, T.J., Nicolas, A. & Coleman, R. G.) 195–205 (Kluwer, Dordrecht, 1991).
- Miyashiro, A. *J. Petrol.* **2**, 277–311 (1961).
- Takahashi, N. & Arai, S. *Sci. Rep. Inst. Geosci. Univ. Tsukuba. Sec. B* **10**, 45–55 (1989).
- Navon, O. & Stolper, E. *J. Geol.* **95**, 285–307 (1987).
- Irvine, T. N. in *Physics of Magmatic Processes* (ed. Hargraves, R. B.) 325–383 (Princeton Univ. Press, 1980).
- Takazawa, E., Frey, A. F., Shimizu, N. & Obata, M. *EOS*, **72**, 537 (1991).
- Takazawa, E., Frey, A. F., Shimizu, N., Obata, M. & Bodinier, J. L. *Nature* **359**, 55–58 (1992).
- Frey, F. A., Shimizu, N., Leinbach, A., Obata, M. & Takazawa, E. *J. Petrol., spec. Lherzolite Issue*, 211–227 (1991).
- Rabinowicz, M., Nicolas, A. & Vigneresse, J. L. *Earth planet. Sci. Lett.* **67**, 97–108 (1984).
- Rabinowicz, M., Ceuleneer, G. & Nicolas, A. *J. geophys. Res.* **92**, 3475–3486 (1987).
- Ceuleneer, G., Nicolas, A. & Boudier, F. *Tectonophysics* **151**, 1–26 (1988).
- Arai, S. *Neues Jb Miner. Monatshefte*, **8**, 347–354 (1987).
- Dick, H. J. B. & Bullen, T. *Contrib. Miner. Petrol.* **86**, 54–76 (1984).

ACKNOWLEDGEMENTS. I thank S. Arai and K. Ozawa for discussions and for reading the manuscript, M. Toriumi and E. Nakamura for suggestions, H. Yoshida for help in microprobe analysis and W. F. McDonough for a review.

Geochemical evidence for melt migration and reaction in the upper mantle

E. Takazawa*, F. A. Frey*, N. Shimizu†, M. Obata‡ & J. L. Bodinier§

* Earth, Atmospheric and Planetary Sciences Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

‡ Department of Geology, Kumamoto University, Kumamoto, Japan 860

§ Centre Géologique et Géophysique-CNRS, Université de Montpellier II, Science et Techniques, Place E. Bataillon, 34095 Montpellier, Cedex 5, France

THE segregation of melts from the Earth's upper mantle into the crust is an important process in the chemical evolution of the crust–mantle system. The processes of melt formation and migration in the upper mantle are inadequately understood, but some important characteristics of these processes can be inferred from upper-mantle rocks exposed at the Earth's surface. The Horoman peridotite body in northern Japan is a layered upper-mantle rock. The major-element compositions of the layers are consistent with their formation as residues from varying extents of melting; however, abundances of rare-earth elements (REE) require additional processes to have occurred¹, such as post-melting enrichment (metasomatism) resulting from reaction with a migrating fluid phase. We report here that chondrite-normalized REE patterns in clinopyroxenes show abrupt changes in slope, which vary with stratigraphic position and rock type. These data can be modelled by chromatographic fractionation as melts migrated through and interacted with peridotite, creating compositional heterogeneities in the upper mantle. In the Horoman peridotite these heterogeneities occur on a scale length of tens of metres.

The Horoman peridotite complex is a fault-bounded tectonic slice emplaced at the southern end of the low-pressure, high-temperature Hidaka metamorphic belt². It consists of cyclic layered sequences of plagioclase lherzolite, lherzolite and

harzburgite with subordinate amounts of dunite, gabbros and pyroxenites (Fig. 1). The peridotite has been divided into two zones²: a lower zone, ~2 km thick, where lithology varies gradationally and layering has a typical wavelength of 100–500 m; and an upper zone, ~1 km thick, characterized by abrupt changes in lithology with thinner layers and more-abundant mafic layers. In addition, Takahashi recognized three distinct rock suites^{3,4}. Our study focuses on the abundant Main Harzburgite–Lherzolite suite in the lower zone.

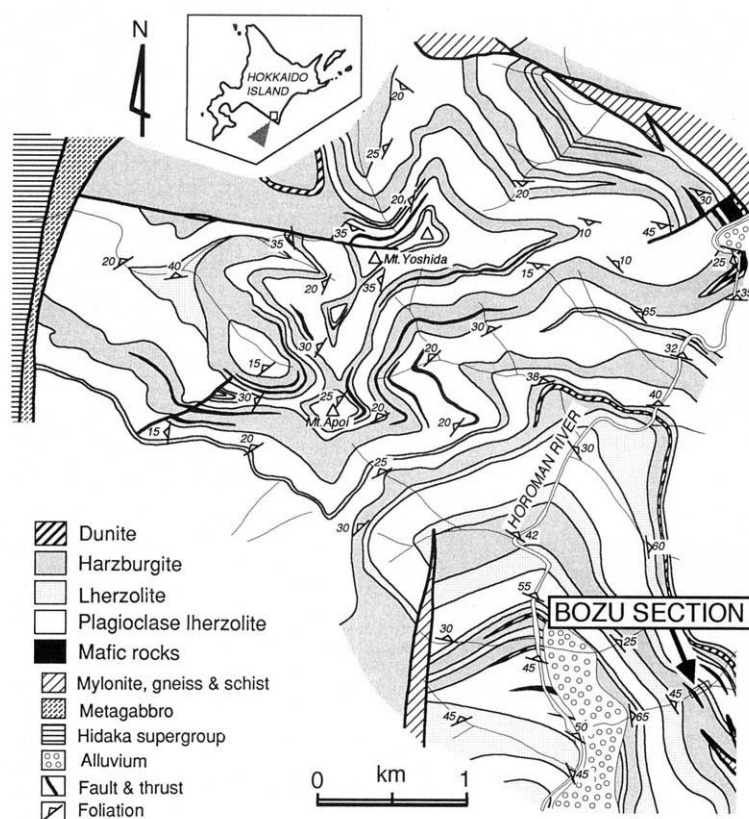
Deformation features are abundant in the peridotites and textures range from porphyroclastic to mylonitic². There is good correlation between modal mineralogy, mineral composition and bulk-rock major element composition^{5,6}: for example, CaO and Al₂O₃ contents decrease in the sequence plagioclase lherzolite to lherzolite to harzburgite. We therefore interpret variations in modal mineralogy as reflecting bulk-rock compositions rather than different pressure–temperature conditions; that is, all of the Horoman peridotites recrystallized at subsolidus conditions corresponding to the plagioclase lherzolite facies. Porphyroclastic pyroxenes with high Al₂O₃ contents² and pyroxene–spinel symplectites^{7,8}, however, provide evidence for earlier high-pressure/high-temperature conditions.

Whole-rock abundances of major elements in Horoman peridotites define simple trends which are consistent with extraction of melts from a source compositionally similar to the plagioclase lherzolite: ~38% MgO, 4% Al₂O₃ and 4% CaO (Fig. 2). The plagioclase lherzolites are similar to mid-ocean-ridge basalt (MORB) in ⁸⁷Sr/⁸⁶Sr (<0.7026) and ¹⁴³Nd/¹⁴⁴Nd (>0.5133)⁹, and they are suitable source rocks for MORB. From previous field, petrographic, petrologic and major element compositional studies of the Horoman peridotite^{1–10} it was concluded that the layering reflects formation and migration of melts in the upper mantle. But abundances of incompatible elements, such as REE, in seven Horoman peridotites from a 250-m section are not consistent with a simple melt-extraction model¹. To understand the process that controlled incompatible element abundances in Horoman peridotites, and perhaps many other peridotites, we studied ~100 m of continuous stratigraphic section exposed in the lower zone of the Horoman peridotite (Fig. 1). This section, composed of a gabbroic layer, harzburgite, lherzolite and plagioclase lherzolite, was sampled in more detail than is typical for geochemical studies of peridotites (Fig. 3).

A wide variety of chondrite-normalized REE patterns occur in this 100-m section (Fig. 4). Clinopyroxenes in the plagioclase lherzolites have REE patterns that are relatively depleted in light REE (LREE), similar to some clinopyroxenes in abyssal peridotites¹². Within 60 m of the gabbro–harzburgite contact, however, REE patterns become convex downwards with abrupt slope changes at Sm or Nd for lherzolites (BZ-140 and BZ-120) and at Dy for harzburgite (BZ-117). Close to the contact with gabbro, harzburgite (BZ-124) has a negative slope without large inflections. The abundances of heavy REE (HREE) in clinopyroxenes decrease from plagioclase lherzolite to lherzolite to harzburgite. In several samples the porphyroclasts (~2 mm) and smaller neoblasts (~0.5 mm) have similar REE contents. Two samples (BZ-147 and BZ-120), however, have core to rim increases in HREE, perhaps indicating that the core of these clinopyroxenes equilibrated with garnet, a mineral which concentrates HREE relative to clinopyroxene. Symplectites (intergrowths of spinel, ortho- and clinopyroxenes) that are interpreted as reflecting a reaction involving the breakdown of garnet^{7,8} provide support for this interpretation.

Because mantle minerals have mineral–melt partition coefficients that yield melting residues with smooth chondrite-normalized REE patterns and lower ratios of LREE to HREE than the source, the data in Fig. 4 are not consistent with the peridotites being residues from varying degrees of melting of a common source. Abrupt slope changes in chondrite-normalized REE patterns could reflect chromatographic effects resulting from percolation of a melt with high LREE/HREE through a

FIG. 1 Geological map of the Horoman peridotite showing the well-developed layering of plagioclase lherzolite, lherzolite and harzburgite. The boundaries of these layers are concordant with foliation planes (indicated, with dip given in degrees). The topographically lower region in the south exposes the lower zone (see text), and the upper zone is exposed at the higher elevations in the north. The layering dips to the north-northeast along the Horoman River and is more nearly horizontal in the upper zone near Mount Apoi and Mount Yoshida. The apparent complexity of the layers on the map results from the topography. This map indicates the occurrence of mafic layers (for clarity their thickness is slightly exaggerated), and it differs from that of Takahashi⁷ which emphasizes relatively thin (<15 m) dunite layers within harzburgite, and earlier maps^{2,5} showing thick dunite layers which we classify as harzburgite. Thin lherzolite layers between harzburgite and plagioclase lherzolite in the upper zone are not shown. The location of the Bozu stratigraphic section (Fig. 3) in the lower zone is indicated. Inset shows the location of the Horoman peridotite in the southern part of central Hokkaido.



porous peridotite matrix¹³. During this process the abundance ratios of geochemically similar elements change considerably because abrupt concentration fronts develop as elements move through the porous matrix at rates that are inversely proportional to their solid-melt partition coefficients. There is also petro-

graphic evidence for fluid migration in the Horoman peridotite: phlogopite and amphibole occur as veins and interstitial grains in some harzburgites and lherzolites¹⁰. But phlogopites have been found only in sample BZ-138 from the Bozu section, and the chondrite-normalized REE pattern for clinopyroxene in this sample is convex upwards, unlike that for clinopyroxene from the phlogopite-free lherzolites. Therefore, there is no obvious connection between the process that formed phlogopite and the process that created unusually high La/Sm.

Bodinier *et al.*¹⁴ and Vasseur *et al.*¹⁵ proposed that the effects of melt percolation through a porous matrix could be modelled by considering melt percolation and diffusion as transport processes with interactions between melt and solid controlled by equilibration at the grain surface followed by volume diffusion in the minerals. We used their approach to interpret the REE variability within the ~100-m stratigraphic section of the Horoman peridotite. The bulk-rock compositions of the plagioclase lherzolites were assumed to be representative of the Horoman peridotite before the percolation process, because they have major element compositions similar to that inferred for primitive mantle¹, coupled with Sr and Nd isotope ratios⁹ and REE abundances typical of MORB-related mantle (Fig. 4 and ref. 1). To explain the decrease in HREE content from plagioclase lherzolite to lherzolite to harzburgite (Fig. 4), we assumed that the matrices interacting with the percolating melt were residua from different degrees of melting of plagioclase lherzolite. Because the mineralogies of the plagioclase lherzolites reflect subsolidus recrystallization⁷, we assumed a spinel lherzolite facies mineralogy for these residua.

To approximate the melt percolation process, we used the following parameters to calculate the results in Fig. 4 (right-hand side): constant mineral proportions and a matrix porosity of 1%; spherical grains of constant radius, 0.5 cm for olivine and 0.1 cm for orthopyroxene and clinopyroxene; a melt velocity of 1 cm yr; mineral-melt partition coefficients from set 1 of ref. 16; a uniform solid diffusion coefficient of 10^{-14} cm² s⁻¹ and a melt diffusion coefficient of 10^{-8} cm² s⁻¹. Although using constant

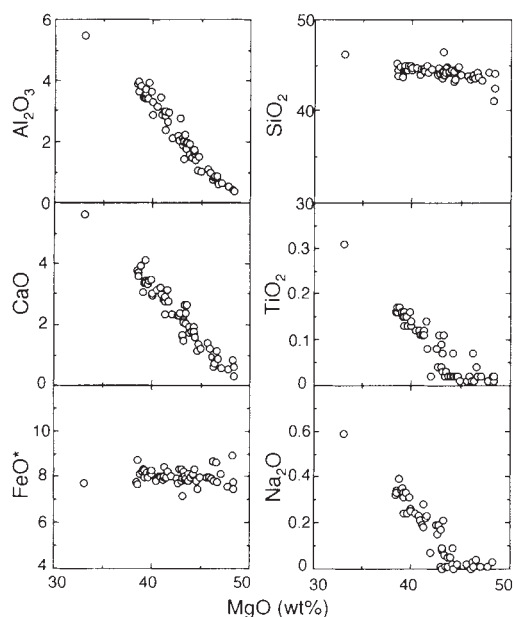


FIG. 2 MgO (wt%) against abundances of other major element oxides (wt%) in 60 Horoman peridotite sample from a 2-km transect along the Horoman River (Fig. 1). All samples in Figs 2–4 are from the Main Harzburgite–Lherzolite suite which is the dominant suite in the Horoman Peridotite^{3,7}. Samples with 3–4% CaO and Al₂O₃ are plagioclase lherzolites; in general lherzolites have lower CaO and Al₂O₃ contents. Data are on an anhydrous basis with all iron reported as FeO (ref. 6).

values is an approximation, these values are physically realistic and consistent with petrography. To match LREE abundances of the Horoman clinopyroxenes, the percolating melt must be relatively enriched in LREE; we assumed a melt with $(La/Yb)_N = 18.7$ (subscript N indicates chondrite-normalized). Because preliminary Sr and Nd isotope data indicate that clinopyroxenes in the lherzolites and harzburgites are isotopically distinct from clinopyroxenes in the plagioclase lherzolite⁹, this introduced melt was not genetically related to the plagioclase lherzolite.

In Fig. 4a and b of the model results, the HREE contents reflect the matrix. This calculated matrix formed as a residue from 15% incremental melting of the plagioclase lherzolite, and it contains clinopyroxene with $(HREE)_N = 3-4$, similar to clinopyroxenes in the lherzolites. The LREE contents of clinopyroxenes reflect interaction with the LREE-rich melt, and the duration of the process. The abrupt slope changes at Nd or Sm observed in the lherzolites develop because LREEs migrate more rapidly than the more compatible HREE. The effect of percolation distance is indicated by the near approach of all REE to equilibrium values at 100 m at time t_2 (Fig. 4a), whereas only La and to a lesser extent Ce have approached equilibrium at a distance of 1 km (Fig. 4b); the concentration fronts of other REE have not yet migrated this distance. To reproduce some aspects of the lherzolite data, such as the zonation of HREE in porphyroclast cores in BZ-120, a more complex model would be required.

Figure 4c and d of the model results shows results for infiltration of the same LREE-rich melt into a matrix formed by 25% melting of plagioclase lherzolite. These calculated results are substantially different because this matrix contains only 0.7% clinopyroxene. Because REEs are more compatible in clinopyroxene than orthopyroxene or olivine, they migrate more rapidly in this clinopyroxene-poor system, and a closer approach to equilibrium is achieved at 100 m and 1 km. The REE patterns

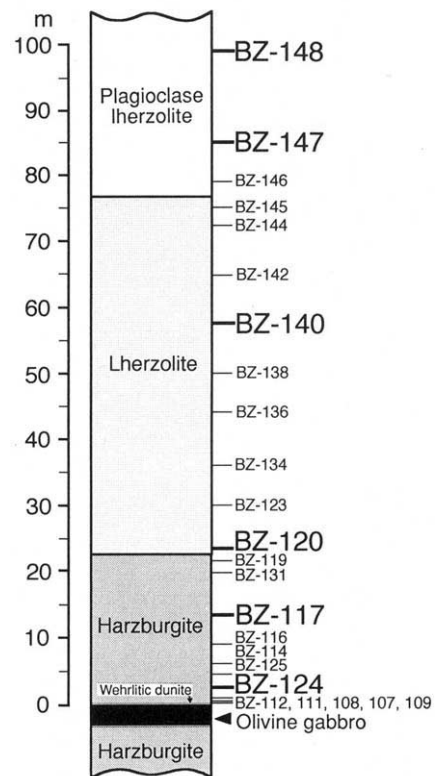
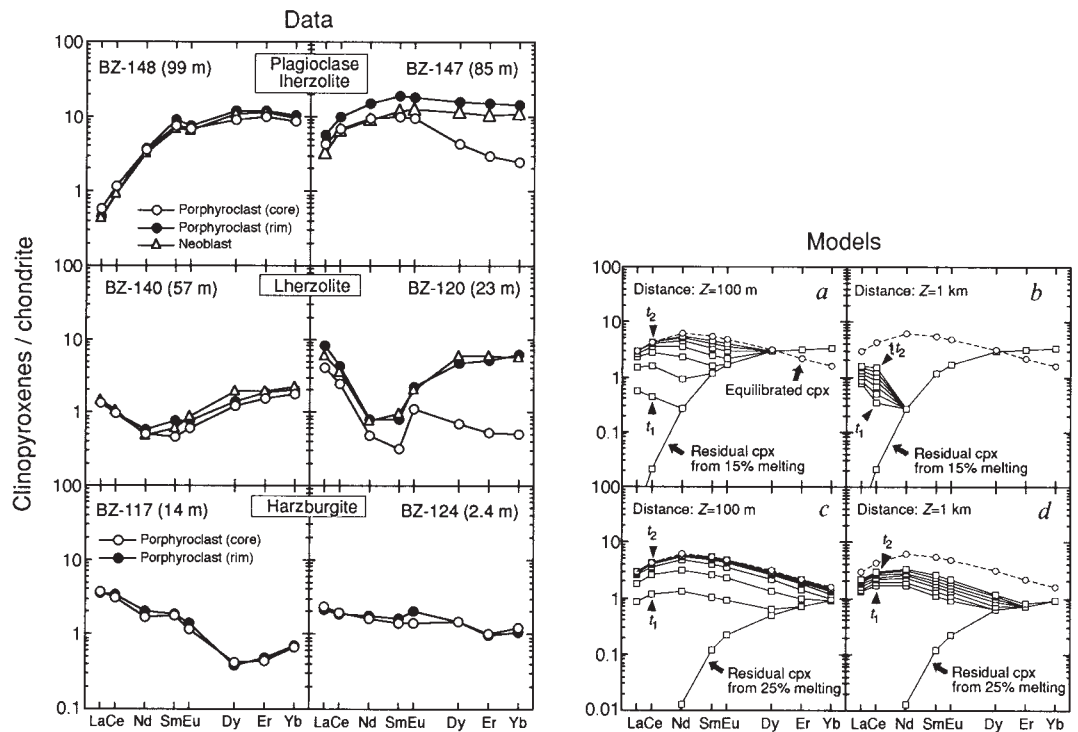


FIG. 3 Sample locations in the Bozu section. Data for samples indicated by large bold type are in Fig. 4. Distance is measured from olivine gabbro-harzburgite contact. The wehrlitic dunite at the gabbro-harzburgite contact is typical of intercalated spinel-rich dunite wherlite and gabbroic layers in the upper zone³.

FIG. 4 Left, chondrite-normalized REE abundances in clinopyroxenes from the Bozu transect; data obtained with the CAMECA 3f ion microprobe at the Woods Hole Oceanographic Institution¹¹. Numbers in parentheses after sample number indicate stratigraphic height in Fig. 3. Right, calculated results for a porous flow model showing the effects of a LREE-rich melt, $(La/Yb)_N = 18.7$, percolating through matrices formed by 15% (top) and 25% (lower) melting. REE contents of these matrices were calculated by incremental melting (1% increments) of a peridotite with the REE content of plagioclase lherzolite 62210 (ref. 1) and the mineralogy of a spinel lherzolite. The chromatographic calculations follow the procedures of Vasseur *et al.*¹⁵ and take into account the effects of advection and diffusion in the melt, and diffusion in the minerals (see text). --○--○, Clinopyroxene in equilibrium with the input melt, $(La/Yb)_N = 18.7$. The initial REE contents of the matrix clinopyroxenes are labelled. —□—□, Transient REE patterns of clinopyroxenes



at various times t ; $t_2 = 20,000$ yr in all panels; $t_1 = 670$ yr in panels a and c and 2,000 yr in panels b and d.

in Fig. 4c and d are qualitatively similar to data for clinopyroxenes in the harzburgites; they have $(\text{HREE})_N \approx 1$ and either no inflection point (BZ-124) or inflections at REEs of higher atomic number, such as Dy (BZ-117).

These model results reflect two processes: formation of residues by incremental melting of plagioclase lherzolite, followed by melt percolation and reaction. In fact, because variations in REE content are correlated with changes in rock type (Fig. 4), we suspect that REE abundance variations and layering resulted from a single process. Changes in porosity and mineral proportions are likely to occur as an exotic melt percolates and reacts with wallrocks. For example, Takahashi^{3,4,7} suggests that the harzburgite layers are residual rocks adjacent to cracks which served as pathways for high-temperature melts, or that the harzburgites formed as reaction products between lherzolite and exotic melts in cracks. We think that our two-step calculation is a reasonable approximation to the process that affected the Horoman peridotites, but this conclusion must be tested with more sophisticated models incorporating space and time variations in porosity and mineral proportions.

A difficulty in using the model results to understand compositional variations in the Bozu Section is that in Fig. 3 distances are measured perpendicular to the gabbroic layer. The exotic melt may not have emanated from this layer. Moreover, this layer may have been a near-vertical dyke. In this case the average path length of the percolating melt is not accurately represented by the measured distance to the gabbro. This complexity may explain the discrepancy between the stratigraphic distances and the longer path lengths required by the model results, and the relatively high La/Sm ratio in lherzolite sample BZ-120 which is close to the lherzolite-harzburgite contact.

Ion exchange chromatographic interactions between a LREE-rich melt and variably depleted peridotite matrix formed as residues from partial melting of the Horoman plagioclase lher-

zolite can qualitatively reproduce features of the unusual REE abundance patterns in clinopyroxenes from the Horoman lherzolites and harzburgites (Fig. 4). Although we would need more realistic models to explain simultaneously the REE abundances and layering, we conclude that chromatographic effects accompanying porous flow were important in controlling the REE geochemistry of the Horoman lherzolites and harzburgites. Mantle xenoliths in alkalic basalts also have a wide range of chondrite-normalized REE patterns; typically patterns with complex shapes and $\text{LREE}/\text{HREE} > 1$ occur in harzburgites¹⁷. Because of their small size (< 1 m), these complexities are not fully understood. We suspect, however, that melt percolation and reaction was also important in controlling the REE content of many mantle xenoliths. □

Received 22 January; accepted 22 July 1992.

1. Frey, F. A., Shimizu, H., Leinbach, A., Obata, M. & Takazawa, E. *J. Petrol. spec. Lherzolites Issue*, 211-227 (1991).
2. Niida, K. *J. Fac. Sci. Hokkaido Univ. Ser. IV*, **21**, 197-250 (1984).
3. Takahashi, N. *J. Min. Petrol. Econ. Geol.*, **86**, 199-215 (1991).
4. Takahashi, N. *Proc. Oman Ophiolite Symp.*, 177-208 (Kluwer, Dordrecht, 1991).
5. Obata, M. & Nagahara, N. *J. geophys. Res.*, **92**, 3467-3474 (1987).
6. Obata, M. & Nishimoto, H. *Kumamoto J. Sci.*, **13**, 25-36 (1992).
7. Takahashi, N. *Nature*, **359**, 52-55 (1992).
8. Takahashi, N. & Arai, S. *Sci. Rep. Inst. Geoscience, Univ. Tsukuba, Sect. B*, **10**, 45-55 (1989).
9. Takazawa, E., Frey, F. A., Shimizu, N., Obata, M. & Nakamura, E. *29th Int. Geol. Congr. Abstr. Kyoto, Japan* (in the press).
10. Arai, S. & Takahashi, N. *Contrib. Mineral. Petrol.*, **101**, 165-175 (1989).
11. Shimizu, N. & Hart, S. R. *A. Rev. Earth planet. Sci.*, **10**, 483-526 (1982).
12. Johnson, K. T., Dick, H. J. B. & Shimizu, N. *J. geophys. Res.*, **95**, 2661-2678 (1990).
13. Navon, O. & Stolper, E. *J. Geol.*, **95**, 285-307 (1987).
14. Bodinier, J. L., Vasseur, G., Vernières, J., Dupuy, C. & Fabries, J. *J. Petrol.*, **31**, 597-628 (1990).
15. Vasseur, G., Vernières, J. & Bodinier, J. L. *J. Petrol. spec. Lherzolites Issue*, 41-54 (1991).
16. Frey, F. A., Green, D. H. & Roy, S. D. *J. Petrol.*, **19**, 462-513 (1978).
17. McDonough, W. F. & Frey, F. A. *Rev. Mineral.*, **21**, 99-145 (1989).

ACKNOWLEDGEMENTS. We thank S. Arai, E. Nakamura, K. Niida and N. Takahashi for discussion, J. Vernières for help with the model calculations and W. F. McDonough for a review. Financial support was provided by NSF and JSPS.

The genomic mutation rate for fitness in *Drosophila*

D. Houle*, D. K. Hoffmaster†, S. Assimacopoulos & B. Charlesworth

Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637-1573, USA

THE mutation rate per genome for local affecting fitness is crucial in theories of the evolution of sex and recombination^{1,2} and of outbreeding mechanisms³. Mutational variation in fitness may also be important in the evolution of mate choice in animals^{2,4,5}. No information is available on the rate at which spontaneous mutations with small effects on fitness arise, although viability (probability of survival to adulthood) has been studied in *Drosophila melanogaster*⁶⁻⁹. These experiments involved the accumulation of spontaneous mutations in the virtual absence of natural selection, in a set of independently maintained lines with a common origin. The rates of decline in mean and increase in variance among lines permit estimation of limits to the mean number of new mutations arising per generation (U) and the average homozygous effect of a new mutation of minor effect (s)^{7,9,10}. For the second chromosome of *D. melanogaster*, the value of U is at least 0.17 (ref. 7), and $(1-h)s$ is less than 0.02, where hs is the average decline in fitness of heterozygotes. As the second chromosome is about 40% of the genome, these data indicate a mutation rate per haploid genome of at least 0.42 for viability. Here we present similar data on the

effects of homozygous spontaneous mutations on a measure of fitness in *D. melanogaster*.

We extracted fifty wild-type second chromosomes from a laboratory population of *D. melanogaster* flies, and assayed them for fitness (see below), viability, longevity and female fecundity. Two hundred copies of the best performing chromosome were maintained in a heterozygous state, and protected from recombination (see legend to Fig. 1) for a total of 44 generations. Each mutation on a wild-type second chromosome, other than a dominant lethal, is effectively neutral. At generations 22, 33 and 44 of mutation accumulation, chromosomes were made homozygous from sets of the lines sampled at random, and were assayed for recessive lethality and sterility. A normal rate of mutability is suggested by the rate at which lethals arise (0.006 lethals per chromosome per generation)⁷.

Forty-seven homozygous viable and fertile wild-type second chromosomes, from mutation-accumulation lines, were randomly chosen at generation 44. Two sublines of each chromosome were established to control for further mutation. The fitness of each chromosome was measured (using a technique modified from ref. 11), by allowing it to compete in a population cage against a marked, homozygous lethal balancer chromosome *In(2LR)bw^{VI}* (ref. 12). Each subline was assayed in two experimental groups. The i th mutation-accumulation chromosome will reach an equilibrium frequency, p_i , determined by the relative fitnesses of the *bw^{VI}/+* and *+/+* genotypes, unless *+/+* has higher fitness. As we measured p_i in adults, we need to make assumptions about the proportion of the total decline in fitness resulting from viability selection (k) to obtain a valid measure of fitness (see Table 1).

In each experimental group, there was a highly significant difference in fitness between line means for mutation-accumulation and control lines (Wilcoxon test: $P = 0.0003$, group 1;

* Present address: Department of Biology, University of Oregon, Eugene, Oregon 97403, USA.

† Present address: Department of Entomology, Michigan State University, East Lansing, Michigan 48824, USA.

TABLE 1 Mutation parameter estimates for fitness

Phenotype	Lines	k	$M \times 10^2$	$V_M \times 10^4$	U (minimum)	$(1-h)s$ (maximum)
Fitness	All lines	0.33	1.28	8.28	0.20	0.065
			1.13-1.44	4.15-12.75	0.13-0.36	0.034-0.096
	Quasi-normal	0.33	1.25	7.09	0.22	0.057
	Quasi-normal	0.58	1.11-1.38	3.27-11.20	0.14-0.43	0.028-0.089
Viability	Quasi-normal ⁶	—	1.04	6.47	0.17	0.062
	Quasi-normal ⁷	—	0.90-1.18	2.57-11.32	0.10-0.37	0.027-0.102
	Quasi-normal ⁸	—	0.38	1.03	0.14	0.027
	Quasi-normal ⁹	—	0.40	0.94	0.17	0.023
			0.17	0.51	0.06	0.030

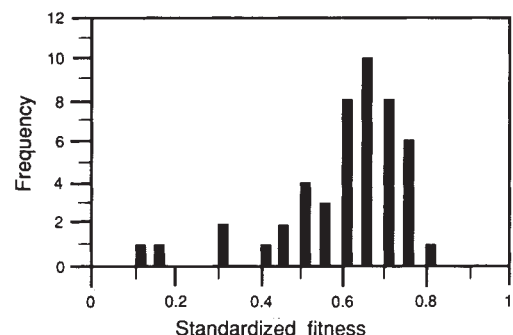
M is the rate of decline per generation of the mean fitness of the mutation-accumulation lines, estimated as the mean fitness in generation 44 of the mutation-accumulation lines relative to the controls, divided by 44. V_M is the rate of increase of the variance in fitness per generation, estimated as $1/44$ of the genetic component of variance of the standardized fitnesses of the mutation accumulation lines. $M = Us$, and $V_M = U(s^2 + V_s)$ where U is the mutation rate per second chromosome per generation, and s and V_s are the mean and variance of selection coefficients against homozygous mutations, respectively^{7,9}. If new mutations have some dominance ($h > 0$), they affect the fitness of heterozygotes, so we estimate $(1-h)s/(1-hs) \approx (1-h)s$ rather than s (ref. 7). Assuming $V_s = 0$, we can obtain limits to U and s from the formulae $U \geq M^2/V_M$ and $(1-h)s \leq V_M/M$. The homozygous fitness of the i th chromosome is given by $w_i = 1 - (1-p_i)/(k_i - 1 + p_i(2-k_i))$. The value of k is unknown, but results for viability make it clear that $k > 0$. If we assume $k=1$, minimizing the decline in fitness, we still obtain an estimate of M twice as large as for viability (see text); therefore $k \ll 1$. More extreme values of k than those shown result in less conservative estimates of U and $(1-h)s$. Variance in k has a minor effect on U and $(1-h)s$, unless it is unrealistically large. Values shown are medians above, and 2.5% and 97.5% quantiles below from 1,000 bootstrap resamplings of the entire data set. Resampling was done hierarchically, first resampling at the level of lines, and then within blocks for each chosen line. Both control and mutation-accumulation lines were resampled.

$P < 0.0001$, group 2). Both means were higher in group 1 than group 2 but the ratio of mutation-accumulation to control mean is similar across-groups, suggesting similar relative effects of mutations (group 1 control mean 0.894, mutation-accumulation mean 0.600; group 2 control mean 0.771, mutation-accumulation mean 0.494; assuming that all selection is on viability $k=1$). This conclusion is corroborated by the lack of significant line-

group interaction for mutation-accumulation lines ($P = 0.34$, 38 and 72 d.f.). Before further analysis, the fitness estimate for each cage was therefore standardized by dividing by the group control mean.

The distribution of line mean fitness (over groups and replicates) is shown in Fig. 1. The distribution is left-skewed with most lines having similar 'quasi-normal' fitness, and a few lines

FIG. 1 Distribution of mean homozygous fitness values for mutation-accumulation lines, assuming $k=1$ (all selection on viability), standardized by the mean fitness of control lines in that group. Fifty second chromosomes were extracted from a laboratory population of *Drosophila melanogaster*, IV. IV was founded with flies from Amherst, Massachusetts in 1975, and subsequently maintained under constant conditions of large population size¹⁹. Chromosomes were extracted using the balancer stock *SM1/Pm*; *spa^{pol}/spa^{pol}*. *SM1* is a balancer chromosome which is multiply inverted to effectively prevent recombination, and marked with *Curly wing* (*Cy*). *spa^{pol}* is a recessive fourth chromosome eye mutant²⁰, used to detect possible contamination of the stocks. The background of X and third chromosomes in the balancer stock was derived from IV. The stock has a *P/R* cytotype, thus avoiding hybrid dysgenesis when crossed to wild-type males¹². Each original second chromosome was assayed for fitness, female fecundity, viability, and adult longevity when homozygous. Two-hundred copies of the single best-performing chromosome have been maintained continuously in a heterozygous state by backcrossing single *SM1/+* males to 3 *SM1/Pm* females each generation. The chromosomes used for fitness measurements were isolated in generation 44 of mutation accumulation. All experiments were carried out at 25 °C on a 12:12 light cycle, using a standard dead yeast medium. The fitness of a homozygous wild-type second chromosome was determined from its equilibrium frequency in competition with the balancer *In(2LR)bw^{v2}* in a population cage (see text for details). We used *bw^{v2}* (ref. 12), because the *Cy* marker carried by the standard second chromosome balancers has incomplete penetrance in our genetic background. Before the measurements of fitness, we placed *bw^{v2}* on the same genetic background as the mutation-accumulation lines, and balanced it over *SM1*. We tested *bw^{v2}* for recombination with a multiply marked test chromosome, and found a rate of recombination of 3% for the entire chromosome (11 recombinants out of 323 backcross progeny). These results suggest that the mutation-accumulation chromosomes probably remain substantially intact on the short timescale of the cage experiments. This is verified by the observed lack of variation in fitness among the control chromosomes compared with the mutation-accumulation lines. Each experimental group assayed both sublines of the mutation-accumulation chromosomes, and two replicated control lines. To start a cage, males from each line were crossed to *SM1/bw^{v2}*; *spa^{pol}/spa^{pol}* females. Ten virgin *bw^{v2}/+* males and females were selected in the next generation, and placed in small plastic population cages which hold six vials of food. Once a week, the two oldest vials in these cages were replaced with new, so that each remained in the cages for three weeks. At the conclusion of each experiment, the adults in each



cage were scored for eye-colour phenotype. A few cages became contaminated with *spa⁺* flies during the course of the experiments, and were discarded. As a control population, we combined the six mutation-accumulation lines with the highest average rank for male and female longevity, and female fecundity early and late in life, measured at generation 33. This new population was maintained in half-pint bottles at large population size. Before and during our experiments, selection should reduce the frequency of any deleterious mutations accumulated on these above-average lines. We used heterozygous samples from this population as controls, as we assume that most of the new mutations are deleterious, which makes this the closest approximation to the fitness of the original homozygous chromosome. Controls for group 1 were taken from this population after one generation of recombination and selection; controls for group 2 after 2 generations. These controls probably lead to a small underestimate of the decline in fitness. In group 1, flies collected from each cage were split so that half were scored by one person and the other half by a different person. There was a significant effect of identity of counter on the ratio of wild-type to brown-eyed (*bw^{v2}*) flies, which was removed statistically before counts were combined to calculate gene frequencies. In group 2, one of the counters of group 1 scored all flies. In group 1, egg samples were removed from each cage every two weeks, and adults scored on days 12 and 20 following collection. Gene frequencies equilibrated about 6 to 8 weeks into the experiment. Group 1 ran for 10 weeks. Given the group 1 results, group 2 was run for only 8 weeks. All cages retained the *bw^{v2}* chromosome. The mean number of flies at the final count was 355.1 per cage.

having greatly inferior fitness. This distribution is similar to that for homozygous viability⁷. The line means were subjected to an outlier test (the Grubbs test)¹³, and two lines that were significant outliers were identified as severely deleterious. These lines probably carry a mutation with a larger effect than most mutations. The parameter estimates should be more reliable when these outliers are excluded⁷. Estimates of mutational parameters are shown in Table 1. U is an underestimate and the selection coefficient, $(1-h)s$, an overestimate if there is variation in selection coefficients among loci^{7,9}. The values of k assumed were 0.58 and 0.33, as analytical and numerical results show that when $k=0.58$, U is minimized and $(1-h)s$ maximized, and when $k=0.33$, the rate of decline in the mean and increase in the variance of viability closely match the results given in ref. 7.

Our experiments provide a measure of line fitness, that includes a much wider variety of fitness components than egg to adult viability, notably female fecundity and male mating ability¹¹. Previous results suggest that the rate of decline in heterozygous fitness is similar to that for heterozygous viability¹⁴, but our estimates of the rate of decline in mean, and of increase in the variance are substantially larger than for viability, regardless of the assumptions made about k . Our best estimate of the genomic mutation rate for fitness U is similar to the rate given in ref. 7 for viability, suggesting that fitness declines more rapidly than viability because mutations that decrease viability also decrease other components of fitness. This interpretation is consistent with other data collected at generations 33 and 44 of mutation accumulation. These data on early and late female fecundity and female and male longevity (D.H., D.K.H., S.A. and B.C., manuscript in preparation), indicate that mutational correlations between these traits are generally strongly positive, and that the traits are subject to similar mutation rates. This also agrees with results of spontaneous mutations affecting development time and viability^{15,16}, and with those for induced variations affecting a variety of traits⁹.

One assumption we make is that the decline in fitness with time is linear on an arithmetic scale. We justify this by the nearly linear decline in viability over a similar number of generations^{9,17}. But the same experiment later showed evidence that viability declined more rapidly, presumably as a result of interactions among mutations. Alternatively, if fitness is determined multiplicatively, the rate of decline in fitness should decrease over time. Both of these departures would bias our estimates towards higher U and lower $(1-h)s$. In contrast, under reasonable assumptions, the variance of s will cause our limits to be several times too conservative. For instance, if the distribution of s is exponential, the per-genome mutation rate would be twice as high as we estimate⁹. Our overall estimates are therefore probably still conservative, in spite of potential counter-biases.

Our results indicate that the mutation rate per haploid genome for fitness in *Drosophila* is at least $0.10/0.4 = 0.25$, and is more likely to be several times larger. Our estimate of mutation rate agrees with those of the earlier *Drosophila* studies, and with indirect estimates for plants¹⁸. It provides substantial empirical support for the high per-genome mutation rate to detrimental alleles postulated in some theories of the evolution of breeding systems^{2,3} and of mate choice^{2,4,5}. □

Received 30 March; accepted 17 July 1992.

1. Crow, J. F. in *Mathematical Topics in Population Genetics* (ed. Kojima, K.-I.) 128-177 (Springer, Berlin, 1970).
2. Kondrashov, A. S. *Nature* **336**, 435-440 (1988).
3. Charlesworth, D. & Charlesworth, B. A. *Rev. Ecol. Syst.* **18**, 237-268 (1987).
4. Charlesworth, B. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Andersson, M. B.) 21-40 (Wiley, Chichester, 1987).
5. Kirkpatrick, M. & Ryan, M. J. *Nature* **350**, 33-38 (1991).
6. Mukai, T. *Genetics* **50**, 1-19 (1964).
7. Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. *Genetics* **72**, 335-355 (1972).
8. Ohnishi, O. *Genetics* **87**, 529-545 (1977).
9. Crow, J. F. & Simmons, M. J. in *The Genetics and Biology of Drosophila*, Vol. 3C (eds Ashburner, M., Carson, H. L. & Thomson, J. N.) 1-35 (Academic, London, 1983).
10. Bateman, A. J. *Int. J. Radiat. Biol.* **2**, 170-180 (1959).

11. Sved, J. A. *Genet. Res., Camb.* **18**, 97-105 (1971).
12. Ashburner, M. *Drosophila. A Laboratory Handbook* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
13. Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, San Francisco, 1981).
14. Simmons, M. J., Preston, C. R. & Engels, W. R. *Genetics* **94**, 467-475 (1980).
15. Mukai, T. & Yamazaki, T. *Genetics* **69**, 385-398 (1971).
16. Yoshimaru, H. & Mukai, T. *Jap. J. Genet.* **60**, 307-334 (1985).
17. Mukai, T. *Genetics* **61**, 749-761 (1969).
18. Charlesworth, B., Charlesworth, D. & Morgan, M. T. *Nature* **347**, 380-382 (1990).
19. Charlesworth, B. & Charlesworth, D. *Heredity* **54**, 71-84 (1985).
20. Lindsley, D. L. & Grell, E. H. *Genetic Variations of Drosophila melanogaster* (Carnegie Institute, Washington, 1968).

ACKNOWLEDGEMENTS. We thank J. T. Ihara, D. Canada, E. Han, J. Houle, A. Pringle, E. T. Norman and K. Uzendowski for their help with experiments. This work was supported by the NIH.

Neural subsystems for object knowledge

John Hart Jr & Barry Gordon*

Department of Neurology, * Department of Psychology, Division of Cognitive Neurology/Neuropsychology and the Epilepsy Center and the Zanvyl Krieger Mind/Brain Institute, the Johns Hopkins University, Baltimore, Maryland 21205, USA

CRITICAL issues in the cognitive neuroscience of language are whether there are multiple systems for the representation of meaning, perhaps organized by processing system (such as vision or language¹⁻⁶), and whether further subsystems are distinguishable within these larger ones. We describe here a patient (K.R.) with cerebral damage whose pattern of acquired deficits offers direct evidence for a major division between visually based and language-based higher-level representations, and for processing subsystems within language. K.R. could not name animals regardless of the type of presentation (auditory or visual), but had no difficulty naming other living things and objects. When asked to describe verbally the physical attributes of animals (for example, 'what colour is an elephant?'), she was strikingly impaired. Nevertheless, she could distinguish the correct physical attributes of animals when they were presented visually (she could distinguish animals that were correctly coloured from those that were not). Her knowledge of other animal properties was completely intact, regardless of input stimulus. To explain this selective deficit, these data mandate the existence of two distinct representations of such properties in normal individuals, one visually based and one language-based. Furthermore, these data establish that knowledge of physical attributes is strictly segregated from knowledge of other properties in the language system.

K.R. was a 70-year-old right-handed retired librarian, with a 10th-grade education (to the age of 16). She developed a subacute neurological illness manifested 12 weeks after onset by impairments in attention and concentration, new learning for verbal and visual material, and naming on visual presentation⁷, found to be specific to the category of animals. This dysnomia and related functions were studied during the next 4 months, with her informed consent. She remained clinically stable during this time, but 3 months after completion of testing she suddenly deteriorated and died.

Box 1 outlines the initial testing used to isolate and characterize her deficit. Her profound deficit in naming, limited to animals (30-60% correct for animals versus 86-99% for other categories), was present despite the input modality (visual or nonverbal sound) or response route (oral or written). Because this deficit was not modality-dependent, it can be attributed to a central processing impairment.

The category-specificity of her dysnomia held even when controlling for frequency, familiarity and visual complexity of the stimuli⁸. Moreover, her naming responses were also quite consistent for each item across multiple administrations (46/61

BOX 1 PERFORMANCE ON NAMING AND RELATED TASKS

(a) Visual perceptual. Performance on these visual perceptual tasks ruled out any deficits in elementary visual perceptual processing.

Extraocular movements intact

No field deficits

Corrected vision = 20/20

No colour blindness

No colour anomia

Intact performance on Poppelreuter figures

Able to assemble cut-up drawings of 8 objects or animals²⁸

(b) General language tasks. Performance on these tasks showed K.R.'s general language functions to be intact with no deficits in her ability to produce animal names.

Intact spontaneous speech²⁹

Intact repetition of simple sentences²⁹

Intact comprehension of two-step commands³⁰

Intact reading and writing of simple sentences²⁹

Able to produce animal names used in the testing in spontaneous speech, in repetition of single words or sentences (26/26 correct), and in oral reading (65/65 correct).

(c) Oral naming of frequency-matched colour photographs. Naming from the animals category was significantly impaired compared to naming from other categories ($\chi^2 = 7.2$, $P < .05$).

Animals	8/20
Food	20/20
Inanimate objects	20/20

(d) Oral naming of colour photographs

Animals	
Animals (general)	10/25 correct
Land animals	5/18
Water animals	2/5
Birds	1/7
Insects	0/6
	18/61 total

Control categories

Occupations/people	12/12
Plants	20/20
Food	25/25
Fruits and vegetables	30/30
Letters	26/26
Numbers	20/20
Colours	20/20
Body parts	17/17
Clothing	24/25
Musical instruments	9/10
Personal items	15/15
Transportation	20/20
Tools	10/10
Household objects	70/70
Toys	20/20
Buildings	10/10
	348/350 total

(e) Category-specific impairment in oral naming of line drawings⁸, matched for frequency, familiarity and visual complexity. Normative data from Snodgrass and Vanderwart⁸. All available items were used. Matching was by either group means or item-to-item, in which case the matching criterion was ± 0.5 s.d. Significance testing was by Chi-squares; expected frequencies were those of K.R.'s performance on nonanimal categories. The same animal pictures were presented on two separate occasions. The comparisons of animal naming against one other category (inanimate objects) showed significant differences; although the comparisons against two categories showed differences of a similar if not greater magnitude, they had smaller *ns*, and failed to reach significance.

Matching by means; three category comparison

Category	<i>n</i>	Mean frequency	Mean familiarity	Mean visual complexity	Number correct	Per cent correct
Animals (1st admin.)	15	11	2.52	3.36	7 ($P < 0.08$)	47%
Animals (2nd admin.)					8 ($P < 0.12$)	53%
Food	15	10	3.02	3.02	14	93%
Inanimate objects	15	11	2.51	3.28	15	100%

Matching by means; two category comparison

Animals (1st admin.)	35	16	2.55	3.65	17 ($P < 0.002$)	49%
Animals (2nd admin.)	35				21 ($P < 0.02$)	60%
Inanimate objects	35	16	2.55	3.62	33	94%

Matching by items; two category comparison

Category	<i>n</i>	Number correct	Per cent correct
Animals (1st admin.)	28	12 ($P < 0.01$)	43%
Animals (2nd admin.)	28	13 ($P < 0.02$)	46%
Inanimate objects	28	24	86%

(f) Oral naming of nonverbal sounds. K.R. was impaired at identifying nonverbal sounds (such as a dog's bark) that were made by or highly associated with the animals that had been presented for naming.

Animal	1/8 (12%)
Objects	7/8 (88%)

(g) Written naming of colour photographs

Animals	11/20
Food	10/10
Objects	10/10

TABLE 1 Property judgments: direct question task (number correct/number administered)

	Physical attributes				Nonphysical attributes and functional properties			
	Colour	4 legs	Not 4 legs	Size	Land/air/sea	Woods/farm/jungle	Food	Pet
Visual words								
Animals	11/30	21/21	14/21	21/30	30/30	41/42	42/42	42/42
Objects	15/15			30/30	30/30			
Auditory words								
Animals	16/34	21/23	7/15	29/37	38/38	38/38	37/38	38/38
		Seeds	Shape					
Fruits and veg.	10/10	10/10	10/10	10/10				
		Wheels						
Vehicles		10/10		10/10	10/10	Carries >10 people 10/10		
Line Drawings								
Animals	2/14*	27/28†	4/13	8/13‡	15/16	16/16	16/16	16/16
		Seeds	Shape					
Fruits and veg.	10/10	10/10	10/10	10/10				
		Wheels						
Vehicles		10/10		10/10	10/10	Carries >10 people 10/10		

These questions were given to four normal age-matched controls. Their performance was intact, except for one error by one normal control in the woods/farm/jungle questions, and one error in the pet questions. Thus, K.R.'s results were not dependent on item selection.

* The colour of the animals was not provided in any of the line drawings.

† The legs were removed from the animal line drawings for these stimuli. The '4 legs' heading refers to the score on animals that had 4 legs and the 'Not 4 legs' heading for those that did not. There is a suggestion from her performance of a tendency for saying animals had 4 legs.

‡ All line drawings were matched to be of roughly the same size in terms of surface area.

items elicited a consistent response on multiple administrations of visual confrontation naming with oral responses), and across response modalities (27/33 items consistent for oral versus written naming)⁹⁻¹¹. K.R. had no overt lexical comprehension disorder, even for the animal names she could not produce¹¹. She scored 100% (65/65) on picture-word matching, with either auditory or visual stimuli, short (5 s) or long (30 s) presentation intervals¹⁰, and with semantically related or semantically and visually related distractors.

K.R.'s residual knowledge of the properties of animals, but not control categories, showed a significant but selective impairment whether tested by direct questioning (such as 'what is the colour of an elephant?') or forced-choice recognition (Tables 1 and 2). She had a deficit for the physical attributes of animals (colour, number of legs and size; for example, answering that an elephant was orange; Table 1), but not for nonphysical attributes of animals, or for any properties of the control categories, including those physical attributes that were impaired for animals. She was not impaired for any functional properties of animals (for example, 'Is an elephant edible? Is it a pet?'),

even for animals she could not name, or describe the physical attributes of, from line drawings.

Cuing of confrontation naming of animal pictures was tested, using auditory or visual word stimuli as cues^{2,10}. Most importantly, cuing with physical attributes (colours or parts, such as 'udder' for a cow) did not facilitate naming (35% versus 40% correct for the uncued condition; $\chi^2 = 0.107$, $P > 0.5$). But cuing with nonvisual perceptual properties ('moo'), or functional properties ('milk') improved K.R.'s naming abilities (95% and 90% and $\chi^2 = 13.80$, $P < 0.001$ and $\chi^2 = 10.99$, $P < 0.001$, respectively). Therefore neither explicit nor implicit tests showed K.R. to have verbal ('language-based') knowledge of the physical attributes of animals, although her knowledge of nonvisual perceptual properties and of other associated functional properties of animals was intact by both types of measures.

K.R.'s deficit did not involve her visually based knowledge of the physical attributes of animals¹²⁻¹⁴, as demonstrated by tests using essentially the same items as the verbal tasks. She was intact in visually matching animal bodies to their respective heads^{13,14} (8/8 correct), pointing to the parts of animals in

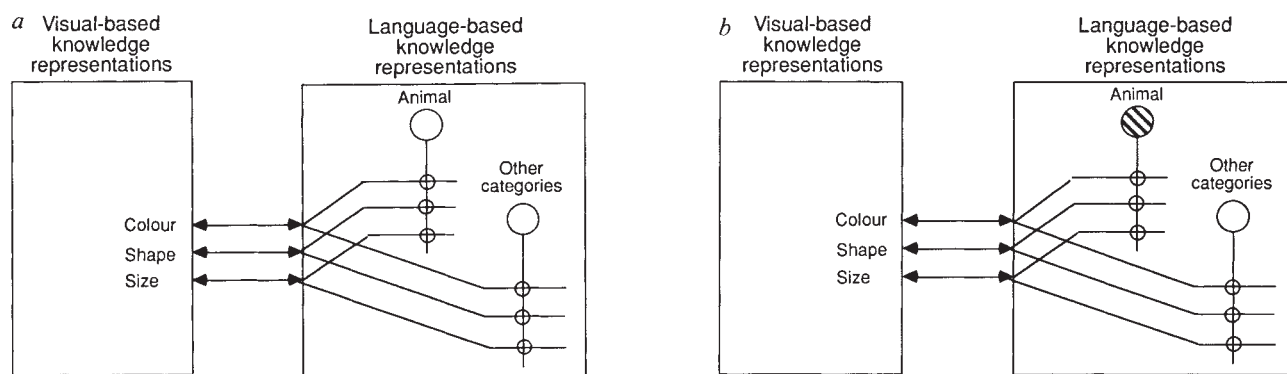


FIG. 1 Hypothetical basis of category specificity in K.R. a, The normal (unlesioned) scheme of knowledge processing. Visual physical attribute information coming from the visual system is divided into many parallel streams for processing in the language system, with the streams organized by semantic categories. The visual physical attribute information inherent

in the language system is similarly organized in these streams. The flow of information in a stream is gated or regulated by activation of a category-level node. b, In K.R.'s case, damage has selectively affected her category-level node for animals, impairing the processing of visual property data that was normally facilitated by that node.

TABLE 2 Property judgments for animals: forced choice task (number correct/number administered)

	Physical attributes				Nonphysical attributes and functional properties			
	Colour	4 legs	Not 4 legs	Size	Land/air/sea	Woods/farm/jungle	Food	Pet
Visual words								
Choice between error on direct question task and correct response	1/19		0/7	0/9				
Choice between correct response on direct question task and distractor error	11/11	21/21	14/14	21/21	15/15	15/15	15/15	15/15
Line drawings								
Choice between error on direct question task and correct response	0/12		0/3	0/5				
Choice between correct response on direct question task and distractor error	2/2	8/8	2/2	8/8	10/10	10/10	10/10	10/10

The task used items on which the patient had made errors on the direct recall task, as well as a subset of the items she had answered correctly. The choices were between her prior errors and the correct answers ('Is an elephant orange or grey?'). The impairment noted on open-ended recall was much more evident on this forced-choice task, although it would have been thought to be easier.

pictures (for an elephant, 'show me the trunk') (27/27 correct)¹⁵, and naming animal body parts from pictures (16/16 correct), even for animals that she could not name. She was also 100% correct (14/14) at discriminating between pictures of animals that showed the proper colour of those animals (grey elephant), versus those pictures that were not colour-appropriate for that particular animal although they were colour-appropriate for animals in general (grey lion). But she could not say which colours were correct for animals that were incorrectly coloured. Clearly, she had no impairment in her internal visual representations of the colours, parts, or general form of animals.

K.R.'s performance obviates several potential problems with interpretation of such deficits. Her deficits must have arisen centrally, because they were present across several input and output modalities. They could not, therefore, be due to task-specific effects. For example, her impairments with the physical attributes of animals were evident whether the task involved recognition, recall, or cueing. Hence they could not be explained by differential dependence on explicit or implicit processing. Finally, the category-specificity of her impairments permits a direct comparison of performance on a task for items from both intact and impaired categories.

K.R.'s pattern of deficits therefore provides unequivocal evidence for at least two distinct representational systems, one language-based and one visually based. Her language-based system must have been responsible for her intact performance with the functional properties of animals, because knowledge of these properties resides only in the language-based system, as described in previous studies^{3,5,12}. Her visually based system must have been responsible for her intact ability to visually recognize the physical attributes of animals, because this information was not in her language system, as the testing showed. No single central representational system that intermingles both types of properties could explain K.R.'s performance.

This demonstration of two distinct representational systems, one subserving visually based knowledge and one for language-based knowledge, is *prima facie* evidence for a distinction already suggested by behavioural and neuroanatomic data^{2,16-18}. But K.R.'s data allow us to go beyond this general specification, claiming that there is, in some cases, a duplication of some types of knowledge for physical attributes across the two systems. It seems apparent that how and in which representational system information is stored depends on both the type of information (particularly the feasibility of forming the appropriate type of code) and individual experience. The visual physical attributes

of objects, as studied in our experiments, must be represented in at least the visual system if they were acquired through visual experience. But they might also be represented in some form in the language system. The congenitally blind may have a reasonably thorough knowledge of visual physical attributes, based entirely in language^{19,20} and acquired exclusively through that system, just as the normally sighted have some knowledge of putatively visual physical attributes in their language system. Some of this information might duplicate that already available visually (albeit in a different code); in other cases, the 'visual' information might be exclusively language-based.

K.R.'s data provide evidence that normally there is a dual representation for visual physical attributes, one visually based and one language-based. This has been suggested by previous studies^{13,21}, but is clearly shown by K.R.'s performance. The best evidence is the dissociation found between K.R.'s intact ability to recognize physical attributes from visually presented pictures, and her impaired ability to verbally describe physical attributes from names, but only for the animal category. K.R.'s performance was consistent with what has been interpreted as a representational deficit rather than an access deficit (degraded representation according to ref. 1)^{2,3,10,11}. Therefore, it would seem that her impairment was a disruption of her language-based representations of the visual physical attributes of animals. Furthermore, we propose that the disruption of her language-based representations also precludes her ability to retrieve this information from the visually based system.

The category-specificity of K.R.'s gap in knowledge of physical attributes is evidence that normally no such gap exists. That is, normally some of the representations of the visual physical attributes of animals can be language-based as well as visually based, and this can be true of some other attributes and items. Alternatively, a less likely explanation for K.R.'s performance is that her verbal representation system was intact, but contained no knowledge of physical attributes and was unable to access this information from the visual representation system, again however, category-specific.

The category-specificity of K.R.'s deficits allows a further conclusion that the language system must contain multiple subdomains of knowledge representation^{1,10,22} (such as visual physical attributes), functionally and, in some sense anatomically, distinguishable. Specifically, K.R. could not identify the colour, size or other visual physical attributes of an animal from its name, but could recall nonvisual or functional properties of the same animal from its name. No deficits in these tasks were seen

with nonanimal categories. It is suggested that category-level information provides a critical gating or facilitative function for property-level knowledge in the language system (Fig. 1).

Relatively little pathological evidence is available for the anatomical basis of category-specific deficits. Most patients have shown temporal lobe pathology^{1,4,12,14,23} but see also ref. 24. Pathology limited to the frontoparietal lobes has been associated with impairments of the nonliving things categories; however, these cases may also have included temporal lobe pathology^{10,22}. K.R.'s brain showed diffuse, mild inflammation attributable to

a paraneoplastic syndrome that involved the cerebral cortex, including both temporal lobes. This was thought to be the basis for her deficits. Based on K.R.'s neuropathology, and those well studied cases in the literature, this type of category-specific deficit involves: (1) left or both temporal lobes; and generally (2) incomplete or patchy disruption, rather than a complete and sweeping disruption. A number of neural processing architectures (hierarchical^{1,10,25} or distributed networks^{22,26,27}) could produce the processing distinctions and anatomical assignments K.R.'s case requires. □

Received 1 April; accepted 22 July 1992.

- Warrington, E. K. & Shallice, T. *Brain* **107**, 829–854 (1984).
- Shallice, T. *The Cognitive Neuropsychology of Language* (eds Coltheart, M., Sartori, G. & Job, R.) 111–127 (Lawrence Erlbaum, Hillsdale, 1987).
- Shallice, T. *Cognitive Neuropsychol.* **5**, 133–142 (1988).
- McCarthy, R. A. & Warrington, E. K. *Nature* **334**, 428–430 (1988).
- Riddoch, M. J., Humphreys, G. W., Coltheart, M. & Funnel, E. *Cognitive Neuropsychol.* **5**, 143–149 (1988).
- Caramazza, A., Hillis, A. E., Rapp, B. C. & Romani, C. *Cognitive Neuropsychol.* **7**, 161–189 (1990).
- Goodglass, H., Kaplan, E. & Weintraub, S. *The Revised Boston Naming Test* (Lea & Febiger, Philadelphia, 1983).
- Snodgrass, J. G. & Vanderwart, M. *J. exp. Psychol. hum. Learn. Mem.* **6**, 174–215 (1980).
- Coughlan, A. K. & Warrington, E. K. *J. Neurol. Neurosurg. Psychiatr.* **44**, 1079–1083 (1981).
- Warrington, E. K. & McCarthy, R. *Brain* **106**, 859–878 (1983).
- Shallice, T. *From Neuropsychology to Mental Structure* 269–306 (Cambridge Univ. Press, Cambridge, 1988).
- Silveri, M. C. & Gainotti, G. *Cognitive Neuropsychol.* **5**, 677–709 (1988).
- Riddoch, M. J. & Humphreys, G. W. *Cognitive Neuropsychol.* **4**, 131–185 (1987).
- Sartori, G. & Job, R. *Cognitive Neuropsychol.* **5**, 105–132 (1988).
- Ogden, J. A. *Brain* **108**, 1009–1022 (1985).
- Paivio, A. *Imagery and Verbal Processes* (Holt, Reinhart & Winston, New York, 1971).
- Beauvois, M. F. *Phil. Trans. R. Soc.* **298**, 35–47 (1982).

- telinde, J. *J. exp. Psychol. Learn. Mem. Cognit.* **8**, 584–598 (1982).
- Zimmer, J. & Keenan, J. M. *J. exp. Psychol. Learn. Mem. Cognit.* **9**, 269–282 (1983).
- Paivio, A. *Mental Representations* (Oxford Univ. Press, New York, 1986).
- Stewart, F., Parkin, A. J. & Hunkin, N. M. *Q. J. exp. Psychol.* **44A**, 261–284 (1992).
- Warrington, E. K. & McCarthy, R. A. *Brain* **110**, 1273–1296 (1987).
- Rapcsak, S. Z., Kaszniak, A. W. & Rubens, A. B. *Neuropsychologia* **27**, 1031–1041 (1989).
- Silveri, M. C., Daniele, A., Giustolisi, L. & Gainotti, G. *Neurology* **41**, 545–546 (1991).
- Rosch, E. *J. exp. Psychol.* **104**, 192–233 (1975).
- Damasio, A. R. *Cognition* **33**, 25–62 (1989).
- Mesulam, M.-M. *Ann. Neurol.* **28**, 597–613 (1990).
- Hart, J. & Gordon, B. *Ann. Neurol.* **27**, 226–231 (1990).
- Goodglass, H. & Kaplan, E. *The Assessment of Aphasia and Related Disorders 2nd edn* (Lea & Febiger, Philadelphia, 1983).
- De Renzi, E. & Vignolo, L. A. *Brain* **85**, 665–678 (1962).

ACKNOWLEDGEMENTS. We thank M. Pettis for assistance in stimulus creation and testing of the patient, P. Talalay for comments on the manuscript, E. Kinch for data analysis, M. McAllister for her artwork and comments on the manuscript, O. A. Selnes for first appreciating the category-specific deficit and for helpful comments and suggestions about the study design, H. Mena at the Armed Forces Institute of Pathology, Washington DC, for the neuropathological assessment and special thanks to J. Hart Sr. This investigation was supported in part by the NIH (the National Institute of Neurological Disorders and Stroke), the Whittier Foundation, the Seaver Foundation and the McDonnell-Pew Program in Cognitive Neuroscience.

Degeneration *in vitro* of post-mitotic neurons overexpressing the Alzheimer amyloid protein precursor

Kazuaki Yoshikawa*†, Takako Aizawa* & Yokichi Hayashi‡

*Department of Molecular Biology, Tokyo Institute of Psychiatry, Setagaya-ku, Tokyo 153, Japan

‡Biochemical Research Laboratory, Morinaga Milk Industry, Zama, Kanagawa 228, Japan

†Present address: Department of Molecular Neurobiology, Tokyo Metropolitan Institute for Neuroscience, Fuchu, Tokyo 183, Japan

A PATHOLOGICAL hallmark of Alzheimer's disease is the deposition of amyloid fibrils in the brain. The principal component of amyloid fibrils is β /A4 amyloid protein^{1,2}, which can be generated by the aberrant processing of a large membrane-bound glycoprotein, the β /A4 amyloid protein precursor (APP)³. To test whether overexpression of APP generates abnormally processed derivatives that affect the viability of neurons, we stably transfected full-length human APP complementary DNA into murine embryonal carcinoma P19 cells. These cells differentiate into post-mitotic neurons and astrocytes after exposure to retinoic acid^{4–6}. When differentiation of the APP cDNA-transfected P19 cells was induced, all neurons showed severe degenerative changes and disappeared within a few days. The degenerating neurons contained large amounts of APP derivatives that were truncated at the amino terminus and encompassed the entire β /A4 domain. These results suggest that post-mitotic neurons are vulnerable to overexpressed APP, which undergoes aberrant processing to generate potentially amyloidogenic fragments.

Mouse embryonal carcinoma cells (P19 cells) are multipotential stem cells which differentiate into a variety of cell types (for example, extra-embryonic endoderm, neurons, muscles and epithelia) in response to chemical agents such as

retinoic acid and dimethyl sulphoxide^{4–7}. We have reported previously that neural differentiation of P19 cells markedly increases the abundance of APP messenger RNA⁸. We first characterized APP molecules expressed in neurally differentiated P19 cells by western blotting (Fig. 1a). APP species with apparent relative molecular masses (M_r) of 105,000–120,000 (105K–120K) markedly increased during days 3–12, but declined thereafter. The time course of altered levels of 105K APP was consistent with that of the change in abundance of mRNA encoding APP695 (ref. 3; Fig. 1b, upper panel). The presence of APP species with M_r 105–120K coincided with the existence of viable neurons in mixed P19 cell cultures during neural differentiation (Fig. 1b, lower panel). On the other hand, APP species with M_r 115–130K were detected on days 15 and 17; these APP species may correspond to the modified forms of APP (APP751/770) expressed in non-neuronal cells, including astrocytes and microglia-like phagocytes^{9,10}.

We then stably transfected APP695 and APP770 cDNAs into P19 cells and isolated four clones for each APP cDNA species (Fig. 1c). The undifferentiated transfectants of APP695 cDNA (APN1–4) had APP immunoreactive bands of $M_r \approx 105K$, and those of APP770 cDNA (APG1–4) at $M_r \approx 115K$. The control transfectants APC1–2 showed only weak bands of APP species at M_r 105–110K. The APP levels of all APP695 cDNA transfectants were higher than those of the APP770 cDNA transfectants.

After treatment with retinoic acid, control transfectants differentiated into mixed populations of neurons and non-neuronal cells (presumably astrocyte precursors), whereas the population of neurons in the APP cDNA transfectants was severely reduced during days 3–5 (Fig. 2a–c). In these cultures, non-neuronal cells showed no degenerative changes and grew to the confluent density on days 8–14 (data not shown). In neuron-enriched cultures (Fig. 2d–f), most of the APP cDNA transfectants had severely degenerated; they had numerous vacuoles within their soma, and many of them had disintegrated neuritic processes. The degenerating neurons in all of the APP cDNA transfectants (APN1–4 and APG1–4) completely disappeared by day 8 (data not shown). Neuronal degeneration in mixed cell cultures was quantified by counting surviving neurons during neural differentiation (Fig. 3). All of the APP cDNA

transfectants tested had much fewer surviving neurons (lower panel) (2–5% of total cells on day 4, less than 1% on day 6) than the control transfectants (31–35% on day 4, and 22–25% on day 6), whereas the number of total viable cells in these APP cDNA transfectants showed no appreciable reduction (upper panel).

We then tested for immunoreactive APP derivatives in total cell extracts of degenerating neurons in neuron-enriched cultures of APP cDNA transfectants (Fig. 4a). Many APP derivatives with M_r 12–100K were detected in extracts from both

APP695 and APP770 cDNA transfectants after exposure to retinoic acid. On the other hand, APP derivatives of M_r 12K and 35–100K were detected in small amounts in extracts of undifferentiated transfectants and transfectants treated with dimethyl sulphoxide, which differentiate into cardiac muscle cells⁷. These derivatives were hardly detected in the control transfectants. Retinoic acid-treated APP cDNA transfectants contained many more APP derivatives with M_r 12–100K than the controls; average relative amounts of APP immunoreactivity determined by densitometric analysis of the western blots were:

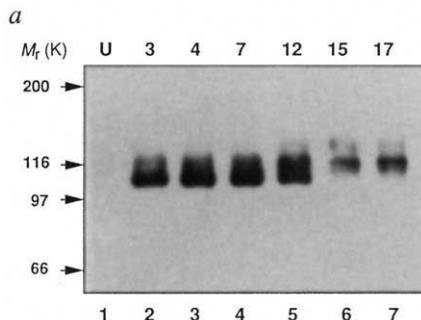
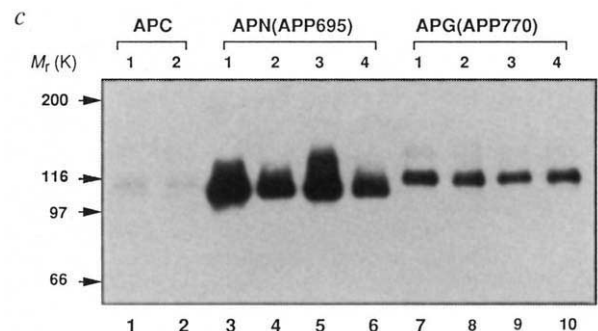
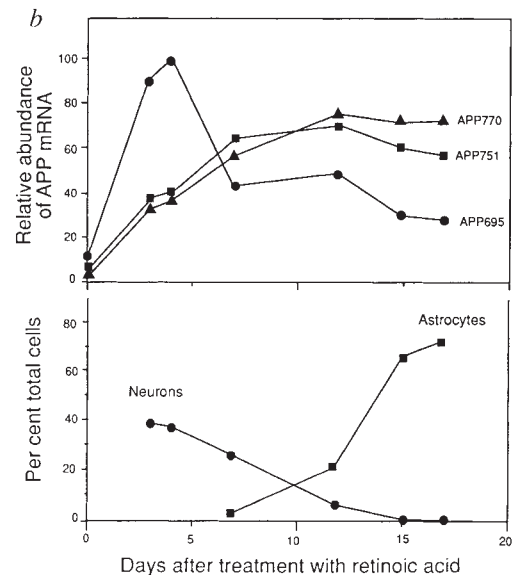


FIG. 1 Expression of APP in P19 EC cells and APP-cDNA transfectants. *a*, Endogenous expression of APP in P19 cells during neural differentiation. P19 cells were cultured for various durations after retinoic acid treatment, and APP in the cell lysates was analysed by immunoblotting. Lane 1, undifferentiated P19 cells (U); lane 2, day 3; lane 3, day 4; lane 4, day 7; lane 5, day 12; lane 6, day 15; lane 7, day 17. *b*, Upper panel, abundances of three species of APP mRNA in P19 cells during neural differentiation. Three APP mRNA species (APP695 mRNA, ●; APP751 mRNA, ■; APP770 mRNA, ▲) were quantified by northern blot analysis using synthetic oligonucleotide probes. Lower panel, proportions of neurons (●) and astrocytes (■) to total cells in mixed cell cultures during neural differentiation. Neurons and astrocytes were identified by phase-contrast micrography and fluorescence immunocytochemistry, respectively. *c*, Expression of APP in APP cDNA-transfected P19 cells. Clones of APP cDNA-transfected cells were cultured under undifferentiated conditions, and collected at a near-confluent density. Lanes 1 and 2, control cells transfected with the vector without APP cDNA inserts (APC1, 2); lanes 3–6, APP695 cDNA transfectants (APN1–4); lanes 7–10, APP770 cDNA transfectants (APG1–4). M_r size markers (in K) are shown on the left.

METHODS. *a*, P19 cells (gift from M. McBurney) were cultured and induced to differentiate as described⁷; P19 cells were aggregated by plating them into bacteriological-grade dishes in the presence of 5×10^{-7} M retinoic acid. After 3-day treatment, aggregates were trypsinized, plated into tissue culture-grade dishes, and cultured in MEM alpha medium (Sigma) supplemented with 10% fetal calf serum (FCS). For immunoblotting, cells were lysed in buffer containing 50 mM Tris-HCl, pH 8.3, 2 mM EDTA and 1 mM phenylmethylsulphonyl fluoride. Lysates (5 μ g protein for each lane) were then electrophoresed on SDS-polyacrylamide gels (8% acrylamide) and transferred to a polyvinylidene difluoride membrane (Immobilon, Millipore) by electroblotting. The membrane was incubated with an antibody raised in rabbits against a synthetic peptide corresponding to the C terminus of APP (residues 671–695 of APP695) (AC1) (1:3,000) and then with peroxidase-labelled goat anti-rabbit IgG(H+L) (1:3,000) (BioRad). Immunoreactive bands were detected by the ECL western blotting detection system (Amersham). Protein concentrations were measured by the bicinchoninic acid method (Pierce). *b*, Three species of APP mRNA were quantified by northern blot analysis using synthetic oligonucleotides complementary to the junctional sequences of mouse APP mRNA¹⁴ as described⁸: total cellular RNA prepared from P19 cells cultured in 100-mm dishes was electrophoresed on a 6% formaldehyde/1.2% agarose gel, and blotted onto a nylon membrane (Hybond-N, Amersham). RNA blots on the membrane were hybridized with the ³²P-labelled oligonucleotides. Densities of autoradiograms were measured with a scanning densitometer. For counting neurons and astrocytes, P19 cells treated with 5×10^{-7} M retinoic acid for 3 days were trypsinized, plated at a density of 5×10^5 cells per 35-mm dish, and cultured in the presence of FCS. Neurons having small soma with long branching processes (distinguishable from non-neuronal cells on day 3 and thereafter) were counted under a phase-contrast microscope; these neurons were positively stained with antibodies against neurofilaments and MAP2 (data not shown). Astrocytes were stained with an antibody against glial fibrillary acidic protein, and the



immunoreactive cells were counted under a fluorescence microscope. Each point represents the mean of 20 measurements. *c*, Full-length human APP695 and APP770 cDNAs^{3,27} were cloned from a human brain cDNA library. After conversion of both ends to *Hind*III sites, the cDNA fragments containing entire coding regions (2.7 and 3.0 kb for APP695 and APP770 cDNA, respectively) were inserted into the unique *Hind*III site of the mammalian expression vector pKGSαN having an SV40 promoter for inserted cDNA²⁸ (gift from T. Nukada). DNA (pKAPN and pKAPG for APP695 and APP770 cDNA, respectively) (10 μ g per 50-mm dish) was transfected into P19 cells by precipitation with calcium phosphate according to the procedure recommended by Stratagene. The transfected cells were incubated in MEM alpha medium supplemented with 10% FCS in the presence of 400 μ g ml⁻¹ G418 (Gibco-BRL). After 14–21 days, about 100 G418-resistant clones for each construct were randomly selected, and propagated in two sets of multiwell plates; one was used for screening by immunostaining with antibody AC1, and the other was kept for further propagation. Clones showing higher APP immunoreactivity were selected and recloned by a limiting dilution method. Four transfected P19 clones for each construct were isolated, and all of the transfected clones constantly expressed higher levels of APP (detected by immunocytochemistry and immunoblotting). Control transfectants (APC1, 2) carrying the vector without cDNA inserts were cloned as already described, except that recloning was omitted. Lysates of the transfectants were electrophoresed on SDS-PAGE (8% acrylamide) and APP detected by immunoblotting.

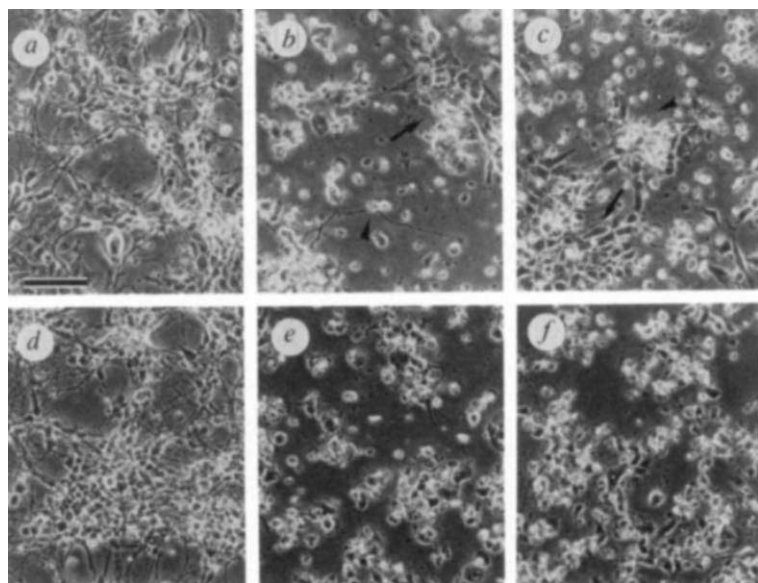


FIG. 2 Phase-contrast micrographs of APP cDNA-transfected P19 cells. Transfectants were aggregated in the presence of 5×10^{-7} M retinoic acid for 3 days, trypsinized, plated at a density of 2×10^6 cells per 100-mm dish, and incubated in alpha MEM medium supplemented with 10% FCS for 4 days (a–c). To enrich the neuronal population, retinoic acid-treated cells were incubated in serum-supplemented medium for 24 h, and then for the next 3 days in a serum-free medium (Opti-Medium 1, Gibco-BRL) (d–f). Cells were cultured in medium containing $200 \mu\text{g ml}^{-1}$ G418. a and d, Control transfectant (APC2); b and e, APP695 cDNA transfectant (APN4); c and f, APP770 cDNA transfectant (APG1). In mixed cell cultures (a–c), neuronal populations in APN4 and APG1 are severely reduced (arrowheads), but large-sized non-neuronal cells grow normally (arrows). In neuron-enriched cultures (d–f), most of the APP cDNA transfectants show severe degeneration. Aggregates having irregular shapes and sizes consist of debris from dead cells (b, c, e, f). Scale bar, $100 \mu\text{m}$ for a–f.

(retinoic acid-treated) APC1-2, 1, APN1-4, 18, and APG1-4, 12; (undifferentiated) APC1-2 was undetectable, APN1-4, 1.1, and APG1-4, 0.8. Most of the APP derivatives in the degenerating neurons were membrane-bound forms (Fig. 4b). Almost identical bands of APP derivatives were detected in two transfectants APN4 and APG1, suggesting that proteolytic processing occurs at specific sites within the APP695 and APP770 molecules. A large amount of the 12K species, presumably the membrane-bound C-terminal fragment generated in the secretory pathway¹¹, was detected in the retinoic acid-treated APP cDNA transfectants. There were many bands corresponding to M_r 16.5–35K; among them, the smallest species (16.5K), which is slightly larger than APP-C100 ($M_r \approx 16$ K; this is the C-terminal fragment of APP and consists of 100 amino-acid residues spanning the β /A4 and cytoplasmic domains¹²), is large enough to contain the entire β /A4 domain. Thus, almost all of the N-terminal-truncated C-terminal fragments, except for the 12K species, are potentially amyloidogenic.

A high abundance of APP mRNA, particularly APP695 mRNA, is expressed in neurons *in vivo* as well as in P19 cell-derived neurons⁸, and endogenous APP in a neuronal cell line is rapidly metabolized by constitutive proteolytic processing¹³. In P19 cell-derived neurons transfected with APP cDNA, overproduction of exogenous APP may exceed the limit of the processing capacity, and subsequently undergo aberrant processing to generate potentially amyloidogenic fragments. As neither undifferentiated P19 stem cells nor differentiated non-neuronal cells show apparent degeneration, post-mitotic neurons might be extremely susceptible to accumulated APP derivatives that could impair some functions indispensable for neuronal survival. Furthermore, the degeneration of neurons may be attributable, at least in part, to the amyloidogenic nature of human APP derivatives expressed in the P19 transfectants; the β /A4 domain of human APP is predicted to form an anti-parallel β -pleated sheet (the basic secondary structure of amyloid fibrils), whereas mouse APP, which is endogenously expressed in differentiated P19 cells, lacks amyloidogenicity owing to amino-acid substitutions in the β /A4 region^{14,15}. Thus, it is possible that exogenous human APP derivatives in the transfectants self-aggregate within the neurons to exert their neurotoxicity.

P19 transfectants overexpressing Ha-ras^{EJ}-1, a human oncogene, differentiate into a normal population of post-mitotic neurons¹⁶. We also observed that P19 cells transfected with cDNA encoding necdin, a brain-specific embryonal carcinoma cell-derived protein¹⁷, differentiated normally into post-mitotic

neurons in the same transfection system as we use here (unpublished results). These results suggest that the degeneration of cDNA-transfected neurons is specific to APP.

Overexpression of the C-terminal fragment of APP (APP-C100) generates amyloid-like fibrils in COS-1 cells¹² and causes neurotoxicity^{18,19}. It is not clear, however, whether APP-C100 or similar fragments are actually generated from full-length APP under pathological conditions such as overexpression²⁰ or mis-sense mutation²¹. Our findings indicate that post-mitotic neurons overexpressing full-length APP can generate potentially amyloidogenic C-terminal fragments. A neuronal origin of β /A4

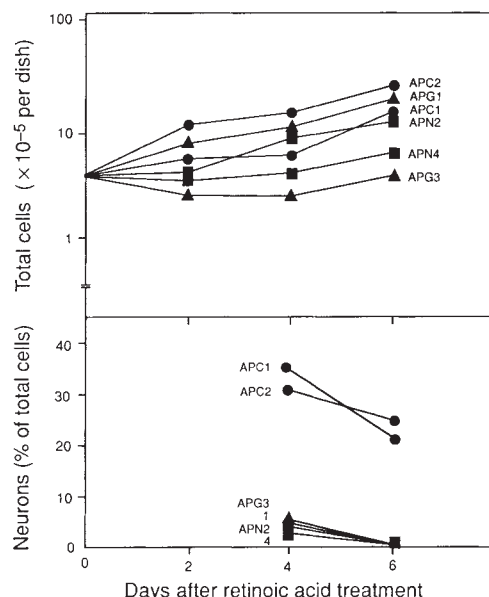
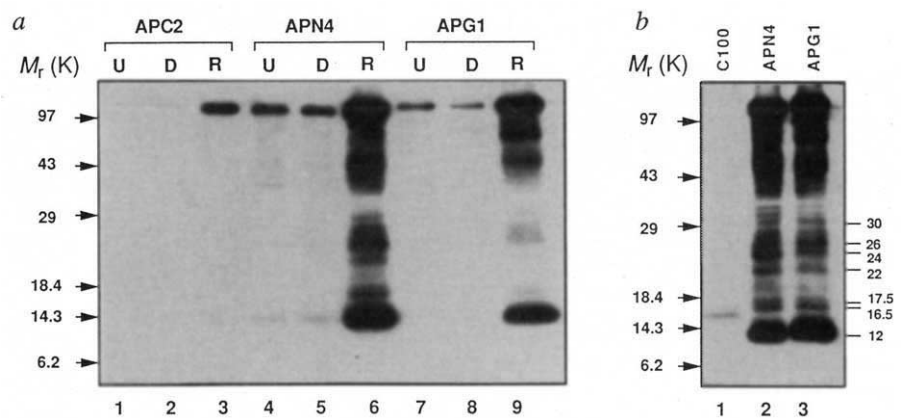


FIG. 3 Neuronal survival of P19 cells transfected with APP cDNA during differentiation. Control transfectants (APC1, 2), APP695 cDNA transfectants (APN2, 4), and APP770 cDNA transfectants (APG1, 3) were treated with retinoic acid for 3 days and then plated at a density of 4×10^5 cells per 35-mm dish. Cells were cultured in serum-supplemented medium containing $200 \mu\text{g ml}^{-1}$ G418. After trypsinizing the cultures on days 2, 4 and 6, the total number of viable cells that excluded trypan blue were counted in a haemocytometer. Neurons were counted as described in Fig. 1b. Surviving neurons in the APP cDNA transfectants on day 6 are less than 1% of total cells.

FIG. 4 APP derivatives in APP cDNA transfectants. *a*, APP derivatives in total cell lysates of transfectants. Transfectants were differentiated by treatment with retinoic acid and dimethylsulphoxide (DMSO). APP in the lysates was analysed by immunoblotting with antibody AC1. Lanes 1–3, control transfectant (APC2); lanes 4–6, APP695 cDNA transfectant (APN4); lanes 7–9, APP770 cDNA transfectant (APG1). Lanes 1, 4 and 7, undifferentiated stem cells (U); lanes 2, 5 and 8, DMSO-treated cells (D); lanes 3, 6 and 9, retinoic acid-treated cells (R). Note the abundant derivatives reacting with the antibody against the C terminus in retinoic acid-treated APP cDNA transfectants. *b*, APP derivatives in the membrane fraction of APP cDNA transfectants. The membrane fraction of the transfectants was prepared, and APP derivatives were detected by immunoblotting with antibody AC1. Lane 1, COS-1 cells transfected with APP-C100 cDNA (C100)/(20 ng protein per lane); lane 2, APP 695 cDNA transfectant (APN4)/(2 µg protein per lane); APP770 cDNA transfectant (APG1)/(4 µg protein per lane). Major bands of both APN4 and APG1 are indicated on the right (M_r in K). The APP-C100 (C100) band corresponds to M_r 16K.

METHODS. *a*, For muscle cell differentiation, control and APP cDNA transfectants were aggregated for 3 days in the presence of 1% DMSO, incubated for 4 days and collected. To prepare neuron-enriched cultures, APP cDNA transfectants were treated with retinoic acid for 3 days, and incubated for 4 days as described in Fig. 2 legend. Enriched neurons in four plates were pooled, centrifuged at 800g for 10 min, and lysed in buffer containing 50 mM



Tris-HCl, pH 8.3, 2 mM EDTA, 1 mM phenylmethylsulphonyl fluoride. Cell lysates were electrophoresed on SDS-PAGE (15% acrylamide), and APP derivatives were detected by immunoblotting with antibody AC1. *b*, To prepare the membrane fraction, cell lysates were resuspended in 20 volumes of lysis buffer, sonicated, and centrifuged at 30,000g for 30 min to collect the pellets as a membrane fraction. COS-1 cells were transfected with the construct of hβ to express APP-C100 as described¹² and the cell lysate was electrophoresed as a molecular size marker. APP derivatives in the membrane fractions of APP cDNA transfectants were analysed by immunoblotting as in *a*.

protein in the extracellular amyloid has been inferred from histochemical and immunohistochemical studies^{22,23}. Moreover, other reports suggest that there is aberrant lysosomal processing of APP (refs 24, 25) and that lysosomal enzymes associate with extracellular amyloid deposits²⁶. Potentially amyloidogenic fragments generated in neuronal lysosomes may thus cause both neuronal degeneration and concomitant deposition into amyloid fibrils at the extracellular space. Stable transfectants overexpressing full-length APP could provide an important system for investigating the APP-associated neuropathological features of Alzheimer's disease and related disorders. □

Received 2 April; accepted 17 July 1992.

- Glennier, G. G. & Wong, C. W. *Biochem. biophys. Res. Commun.* **122**, 1131–1135 (1984).
- Masters, C. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4245–4249 (1985).
- Kang, J. et al. *Nature* **325**, 733–736 (1987).
- McBurney, M. W., Jones-Villeneuve, E. M. V., Edwards, M. K. S. & Anderson, P. J. *Nature* **299**, 165–167 (1982).
- Jones-Villeneuve, E. M. V., McBurney, M. W., Rogers, K. A. & Kalnins, V. I. *J. Cell Biol.* **94**, 253–262 (1982).
- McBurney, M. W. et al. *J. Neurosci.* **8**, 1063–1073 (1988).
- Rudnicki, M. A. & McBurney, M. W. In *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 19–49 (IRL, Oxford, 1987).
- Yoshikawa, K., Aizawa, T. & Maruyama, K. *Biochem. biophys. Res. Commun.* **171**, 294–309 (1990).
- Haass, C., Hung, A. Y. & Selkoe, D. J. *J. Neurosci.* **11**, 3783–3793 (1991).
- Aizawa, T., Haga, S. & Yoshikawa, K. *Dev. Brain Res.* **59**, 89–97 (1991).
- Esch, F. S. et al. *Science* **248**, 1122–1124 (1990).
- Maruyama, K., Terakado, K., Usami, M. & Yoshikawa, K. *Nature* **347**, 566–569 (1990).
- Weidemann, A. et al. *Cell* **57**, 115–126 (1989).
- Yamada, T., Sasaki, H., Dohura, K., Goto, I. & Sakaki, Y. *Biochem. biophys. Res. Commun.* **158**, 906–912 (1989).
- Johnstone, E. M., Chaney, M. O., Norris, F. H., Pascual, R. & Little, S. P. *Molec. Brain Res.* **10**, 299–305 (1991).
- Bell, J. C., Jardine, K. & McBurney, M. W. *Molec. cell. Biol.* **6**, 617–625 (1986).
- Maruyama, K., Usami, M., Aizawa, T. & Yoshikawa, K. *Biochem. biophys. Res. Commun.* **178**, 291–296 (1991).
- Yankner, B. A. et al. *Science* **245**, 417–420 (1989).
- Fukuchi, K. et al. *Biochem. biophys. Res. Commun.* **182**, 165–173 (1992).
- Palmert, M. R. et al. *Science* **241**, 1080–1084 (1988).
- Goate, A. et al. *Nature* **349**, 704–706 (1991).
- Allsop, D. et al. *Neuropath. appl. Neurobiol.* **15**, 531–542 (1989).
- Probst, A., Langui, D., Ipsen, S., Robakis, N. & Ulrich, J. *Acta neuropathol.* **83**, 21–29 (1991).
- Benowitz, L. et al. *Expl. Neurol.* **106**, 237–250 (1989).
- Goide, T. E., Estus, S., Yourkin, L. H., Selkoe, D. J. & Yourkin, S. G. *Science* **255**, 728–730 (1992).
- Cataldo, A. M. & Nixon, R. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3861–3865 (1990).
- Kitaguchi, N., Takahashi, Y., Tokushima, Y., Shiojiri, S. & Ito, H. *Nature* **331**, 530–532 (1988).
- Nukada, T., Mishina, M. & Numa, S. *FEBS Lett.* **211**, 5–9 (1987).

ACKNOWLEDGEMENTS. We thank K. Maruyama, F. Kametani and M. Usami for advice and discussion, D. Allsop for critical review of the manuscript, K. Kashiwagi and K. Takagi for technical assistance, and K. Kato for photographic assistance. This work was supported by the Ministry of Health and Welfare of Japan (K.Y.) and the Ministry of Education, Science and Culture of Japan (K.Y.).

Antisense c-myb oligonucleotides inhibit intimal arterial smooth muscle cell accumulation *in vivo*

Michael Simons*†, Elazer R. Edelman‡§, Jean-Luc DeKeyser§, Robert Langer§ & Robert D. Rosenberg*†¶

* Department of Biology, † Division of Health, Science and Technology, and § Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA and ‡ Department of Medicine, Harvard Medical School, Beth Israel Hospital, Boston, Massachusetts 02122, USA

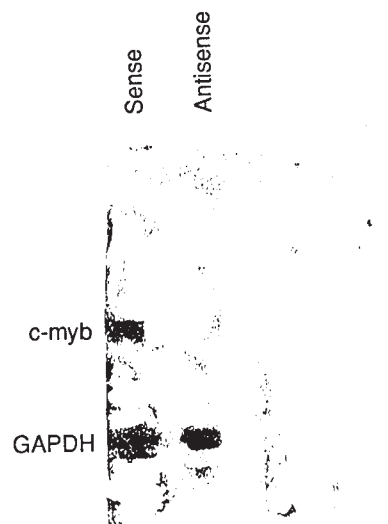
SYNTHETIC antisense oligonucleotides have been used to dissect gene function *in vitro*. Technical difficulties prevented the use of this approach for investigating the effect of gene products *in vivo*. Here we report the use of local delivery of antisense c-myb oligonucleotide to suppress intimal accumulation of rat carotid arterial smooth muscle cells. Our results suggest that antisense oligonucleotides can be used to define the *in vivo* biological role of specific macromolecules in the blood vessel wall and could potentially serve as a new class of therapeutic agents for cardiovascular disorders.

The proto-oncogene *c-myb* is homologous to the transforming gene product of the avian myeloblastosis virus and was originally thought to be present only in haematopoietic cells^{1,2}. But *c-myb* and its various family members are synthesized in diverse cell types and regulate cell growth and differentiation³. The proto-oncogene is involved in mitogen-induced proliferation of vascular smooth muscle cells (SMC) which constitutes a major pathway of atherogenesis⁴. Previous studies demonstrated that *in vitro* stimulation of SMC growth is followed by an elevation in *c-myb* messenger RNA levels and suppression of the message increase by antisense oligonucleotide halts cell proliferation^{5,6}.

We have used a rat carotid artery injury model to show that

¶ To whom correspondence should be addressed.

FIG. 1 Northern blot analysis of carotid artery *c-myb* mRNA levels. The levels of *c-myb* mRNA in the left carotid artery of sense (left lane) or antisense (right) oligonucleotide-treated rats. The left carotid artery in the neck was removed 2 weeks after balloon angioplasty from four rats treated with F127 pluronic gels containing 200 μ g of sense or antisense oligonucleotide. The two antisense- and two sense-treated segments exposed to gel were individually pooled, frozen in liquid nitrogen, crushed, and total RNA was isolated in parallel from all animals as previously described²³. Subsequently, poly(A)⁺RNA was prepared in parallel from all tissues using Invitrogen Micro-Fast Track columns (Invitrogen, San Diego, CA). Ethidium stained gels showed no degradation of RNA samples as evidenced by intact ribosomal bands and mRNA smears centred at a molecular size of about 3–4 kilobases (kb). Samples of 10 μ g of poly(A)⁺RNA were subjected to electrophoresis on a 1% formaldehyde-agarose gel, transferred to a Gene Screen Plus membrane (New England Nuclear, Boston, MA) and hybridized with cDNA probes for *c-myb* and GAPDH as previously described⁵. The c.p.m. used in *c-myb* hybridization were 10 times those used for GAPDH hybridization. The blot was then exposed using a Betascope 603 Blot Analyzer (Betagen, Waltham, MA). The counts in the GAPDH band of the sense-treated sample were 6.5 times the counts in the *c-myb* band of the same sample, and the counts in the *c-myb* band of the antisense-treated sample were at background levels. The counts in the GAPDH band of the sense-treated sample were 35% greater than the counts in the GAPDH band of the antisense-treated sample. On the basis of the available data, the levels of *c-myb* and GAPDH mRNA in the sense-treated sample differ by about 65-fold assuming equivalent specific activity of cDNA probes, and the levels of *c-myb* mRNA in the antisense-treated animals must be less than 5% of the sense-treated controls.

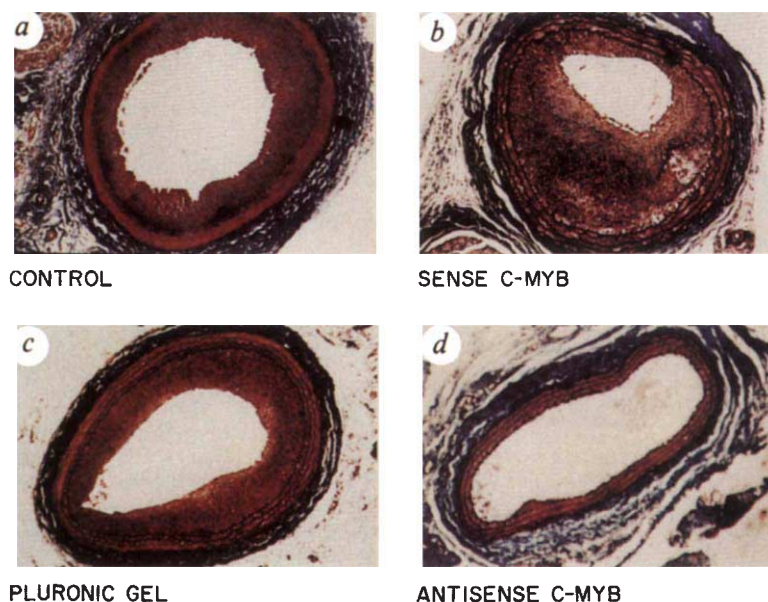


local delivery of antisense *c-myb* phosphorothioate oligonucleotide suppresses SMC migration/growth *in vivo*. To this end, the left common carotid arteries of rats were subjected to balloon angioplasty which denudes endothelium and induces a highly reproducible intimal migration/proliferation of SMC over the entire length of the affected blood vessel^{7,8}. The antisense oligonucleotide (GTGTCTGGGGTCTCCGGGC) (position 4 to 22 of the mouse *c-myb* sequence) or the corresponding sense oligonucleotide (GCCCGGAGACCCCGACAC) were added to pluronic solutions at 1 mg ml⁻¹ and kept at 4 °C. Immediately after balloon angioplasty, 200 μ l of solution was applied to the carotid artery from which the adventitia was stripped. The treated area constitutes about half the blood vessel and represents the portion that lies within the neck. On contact with tissues at 39 °C, the solution gelled instantaneously generating a translu-

cent layer that enveloped the treated region. The wounds were closed immediately after application of gel, and the rats were returned to their cages. Inspection of additional animals revealed that pluronic gel disappears over 1–2 h.

We initially ascertained the effect of *c-myb* oligonucleotide in suppressing oncogene mRNA levels in the rat carotid artery 2 weeks after injury when the extent of SMC accumulation had reached a maximum. The treated portion of the blood vessel was surgically removed from five pairs of antisense and sense-treated rats. Northern blot analysis of one set of animals (Fig. 1) shows that injured carotid artery treated with antisense oligonucleotide exhibits no detectable levels of *c-myb* mRNA whereas injured carotid artery treated with sense oligonucleotide possesses significant amounts of oncogene mRNA. Examination of the injured carotid arteries of three additional sets of animals

FIG. 2 Effect of antisense and sense *c-myb* oligonucleotides on neointimal formation in rat carotid arteries subjected to balloon angioplasty. Representative cross-sections from the carotid artery of an untreated rat (a), a rat treated with pluronic gel containing 200 μ g sense oligonucleotide (b), a rat treated with pluronic gel (c), and a rat treated with pluronic gel containing 200 μ g antisense oligonucleotide (d). (Mason trichrome, magnification $\times 80$.) Male Sprague-Dawley rats (average weight 500 g) were anaesthetized with Nembutal (4 mg per 100 g), and the left carotid artery of each animal was isolated by a midline cervical incision, suspended on ties and stripped of adventitia as previously described²⁴. A 2F Fogarty catheter was introduced through the external carotid artery of each rat, advanced to the aortic arch, the balloon was inflated to produce moderate resistance to catheter movement and then gradually withdrawn to the entry point. The entire procedure was repeated three times for each animal. The oligonucleotides were added at a concentration of 1 mg ml⁻¹ to 25% (w/v) solutions of F127 pluronic gels prepared as outlined by the manufacturer (BASF Wyandotte Corporation, Wyandotte, MI) and maintained at 4 °C. Prechilled pipettes and tips were used to apply the solution to the left carotid artery of each animal with generation of a gel surrounding the exposed blood vessel, and the wounds were then closed. At the time of killing (2 weeks later), the animals were anaesthetized with Nembutal and perfused with 150 cc normal saline under a pressure of 120 mm Hg. The carotid arteries were removed, fixed in 3% formalin, and processed for light microscopy in a standard manner.



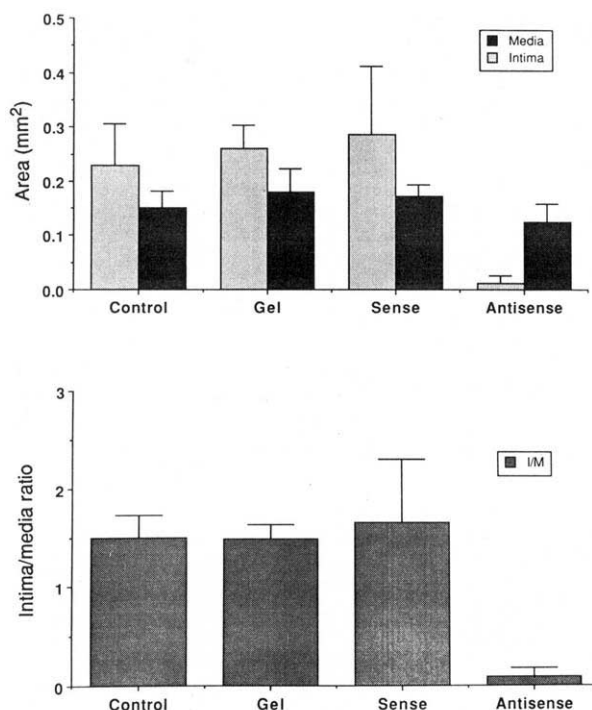


FIG. 3 Intimal and medial cross-sectional areas, as well as intimal/media ratios of untreated and treated rat carotid arteries. The upper panel shows mean cross-sectional areas of the intimal (light bars) and medial (dark bars) regions of rat carotid arteries which were untreated, treated with pluronic gel, treated with pluronic gel containing 200 μ g sense oligonucleotide, and pluronic gel containing 200 μ g antisense oligonucleotide. The lower panel depicts the ratio of intimal to medial areas for the same four groups of animals. The data are provided as means $\pm$ 1 standard deviation.

showed identical results. Analysis of one other set of animals revealed a minimal amount of oncogene mRNA in the antisense oligonucleotide-treated artery as compared with that exposed to sense oligonucleotide. The equivalent loadings of mRNA were documented by hybridization against a glyceraldehyde phosphate dehydrogenase (GAPDH) complementary DNA probe and examination of ethidium bromide-stained gels.

We then determined the effect of antisense oligonucleotide on intimal SMC accumulation 2 weeks after angioplasty in 22

rats who were treated with pluronic gels containing antisense oligonucleotide ($n=8$); sense oligonucleotide ($n=5$); no oligonucleotide ($n=4$); or were untreated ($n=5$). Morphological examination revealed that minimal intimal accumulation occurred with antisense oligonucleotide whereas extensive SMC accumulation was observed in the other treatment or control regimens (Fig. 2). The measurements of arterial segments document suppression of intimal SMC accumulation by antisense oligonucleotide without any apparent effect on medial SMC viability (Fig. 3). We also did experiments using an antisense two-base-pair mismatch oligonucleotide (GTGCCGGGGTCT-TCGGGC) to substantiate further the specificity of the antisense oligonucleotide. Previous *in vitro* results show that antisense mismatch oligonucleotide does not reduce *c-myc* mRNA levels or SMC growth⁵. The *in vivo* data reveal that antisense mismatch oligonucleotide is unable to inhibit intimal SMC accumulation (the ratios of intimal to medial cross-sectional areas for mismatch oligonucleotide, sense oligonucleotide, and antisense oligonucleotide are 1.05 ± 0.14 ($n=3$), 1.02 ± 0.07 ($n=2$), and 0.16 ± 0.03 ($n=2$), respectively). We have also used local single bolus intravascular delivery of oligonucleotides in rabbits during catheter-induced injury of iliac arteries and angiographically demonstrated significant reduction of SMC growth for up to 2 months after treatment.

Rat carotid arteries were examined to evaluate the lengthwise extent of suppression of SMC accumulation. Figure 4 shows a treated region of a rat carotid artery in the neck and an untreated segment of the same blood vessel in the chest. Similar patterns of SMC accumulation were noted with other antisense oligonucleotide-treated rats. The action of antisense oligonucleotide is apparently limited to the portion of carotid artery immediately in contact with gel. This supposition was confirmed in two additional rats where the extent of gel deposition was marked with silk ties which reveal a transition from minimal to extensive intimal accumulation over 3–4 mm of arterial length just proximal to application of gel (Fig. 4).

The injury of the arterial wall induces synthesis and/or release of numerous mitogens which bind to their own specific receptors eventually leading to migration/proliferation of SMC^{9–12}. Recent attempts to inhibit this event used prolonged infusions of specific antibodies against individual mitogens, but were only partially successful, or had no effect^{13,14}. An alternative strategy for suppressing the above process would be to inhibit production of a key intracellular intermediate of a common mitogenic pathway required by all growth factors. Data summarized in this communication indicate that *c-myc* expression is essential

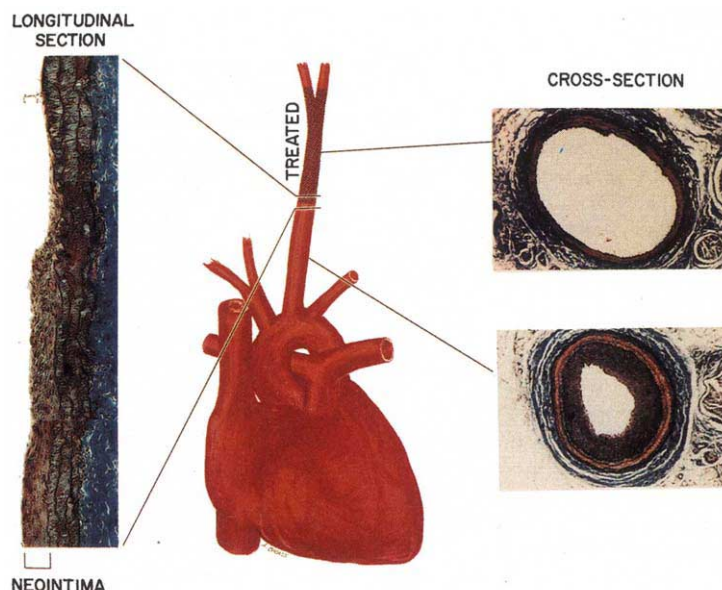


FIG. 4 Spatial distribution of the antiproliferative effect of antisense oligonucleotides. Typical cross-sections (Mason trichrome, $\times 80$) from treated and untreated segments of the left carotid artery of a rat are shown on the right and a longitudinal section of the transitional area between these zones (Mason trichrome, $\times 200$) is depicted on the left. The shaded area represents the extent of application of pluronic gel.

for the function of this common mitogenic pathway. We have previously demonstrated that the drug heparin and blood vessel wall heparan sulphates probably also inhibit proliferation of SMC by suppression of *c-myc*¹⁵. Earlier studies have revealed that suppression of SMC migration/growth for two to three days after arterial injury is sufficient to effect cell proliferation at 2 weeks¹⁶. Therefore, the action of antisense oligonucleotide could have been limited to a few days. But preliminary experiments with radioactively labelled phosphorothioate oligonucleotide indicate that the amounts of this component in the blood vessel wall remains roughly constant from 90 min to at least 72 h. The stability of phosphorothioate oligonucleotide to nuclease action¹⁷ suggests that the effects of these agents may be manifest for a much longer time.

To the best of our knowledge, these studies constitute the first reported use of antisense oligonucleotide to inhibit synthesis of a normal gene product under *in vivo* conditions with a sub-

sequent effect on a cellular process. Our positive results are probably due to local delivery of high concentrations of an antisense oligonucleotide modified to reduce *in vivo* degradation. The success of this strategy suggests that antisense oligonucleotides can be used *in vivo* to suppress other low abundance gene products segmentally in the blood vessel wall with subsequent determination of the resultant physiological effects. Furthermore, intimal SMC accumulation constitutes a major pathological event responsible for long-term failure of coronary and peripheral arterial bypass grafts as well as the development of stenosis following coronary artery angioplasty^{18,19}. Several promising approaches have been described for suppressing this process, but none have proved clinically useful as yet²⁰⁻²². The antisense oligonucleotides constitute a new class of therapeutic agents which can be locally targeted to inhibit specific gene products that result in suppression of intimal SMC accumulation, and possibly other acute pathologic processes. □

Received 15 May; accepted 30 July 1992.

1. Luscher, B. & Eisenman, R. N. *Genes Dev.* **4**, 2235-2241 (1990).
2. Anfossi, G., Gewirtz, A. M. & Calabretta, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3379-3383 (1989).
3. Thompson, C. B., Challenor, P. B., Neiman, P. E. & Groudine, M. *Nature* **319**, 374-380 (1986).
4. Ross, R. *N. Engl. J. Med.* **314**, 488-500 (1986).
5. Simons, M. & Rosenberg, R. D. *Circ. Res.* **70**, 835-843 (1992).
6. Brown, K. E., Kindy, M. S. & Sonenshein, G. E. *J. biol. Chem.* **267**, 4625-4630 (1992).
7. Clowes, A. W., Reidy, M. A. & Clowes, M. M. *Lab. Invest.* **49**, 327-333 (1983).
8. Edelman, E. R., Adams, D. H. & Karnovsky, M. S. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3773-3777 (1990).
9. Deuel, T. F. & Huang, J. S. *J. clin. Invest.* **74**, 669-676 (1984).
10. Linder, V., Lappi, D. A., Baird, A., Majack, R. A. & Reidy, M. A. *Circ. Res.* **68**, 106-113 (1991).
11. Jawien, A., Bowen-Pope, D. F., Lindner, V., Schwartz, S. M. & Clowes, A. M. *J. clin. Invest.* **89**, 507-511 (1992).

12. Hansson, G. K., Jonasson, L., Holm, J., Clowes, M. M. & Clowes, A. W. *Circ. Res.* **63**, 712-719 (1988).
13. Ferns, G. A. A. *et al. Science* **253**, 1129-1132 (1991).
14. Lindner, V. & Reidy, M. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3739-3743 (1991).
15. Reilly, C. F., Kindy, M. S., Brown, K. E., Rosenberg, R. D. & Sonenshein, G. E. *J. biol. Chem.* **264**, 6990-6995 (1989).
16. Guyton, J. R., Rosenberg, R. D., Clowes, A. W. & Karnovsky, M. J. *Circ. Res.* **46**, 625-634 (1980).
17. Agrawal, S., Tamsamani, J. & Tang, J. Y. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7595-7599 (1991).
18. Dilley, R. J., McGeachie, J. K. & Prendergast, F. J. *Atherosclerosis* **63**, 99-107 (1987).
19. Nobuyoshi, M. *et al. J. Am. Coll. Cardiol.* **17**, 433-439 (1991).
20. Jonasson, L., Holm, J. & Hansson, G. K. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2303-2306 (1988).
21. Powell, J. S. *Science* **245**, 186-189 (1989).
22. Pukac, L. A. *Am. J. Path.* **139**, 1501-1509 (1991).
23. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156-159 (1987).
24. Edelman, E. R., Nugent, M. A., Smith, L. T. & Karnovsky, M. J. *J. clin. Invest.* **89**, 465-473 (1992).

Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C

Randy Legerski & Carolyn Peterson

Department of Molecular Genetics, The University of Texas, M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, USA

XERODERMA pigmentosum (XP) is a rare human autosomal recessive disease characterized by solar sensitivity, high predisposition for developing cancers on areas exposed to sunlight, and, in some cases, neurological abnormalities^{1,2}. XP cells are defective in DNA repair³, and complementation of this defect has been used to identify eight genetic groups (A-G and variant)⁴. We have developed a simple, highly efficient complementary DNA expression system for use in human cells⁵. Here we use this system to isolate a cDNA clone that restores the ultraviolet sensitivity and unscheduled DNA synthesis of XP-C cells to normal levels. The XP-C complementing clone XPCC encodes a highly hydrophilic protein which is composed of a predicted 823 amino acids and shares limited homology with the product of the yeast DNA repair gene *RAD4*. The XPCC transcript is undetectable by northern blotting in most XP-C cell lines examined.

The plasmid pEBS7 is an Epstein-Barr virus-based cDNA expression vector that replicates extrachromosomally in human cells⁵. A library containing over 20 million independent clones was prepared in pEBS7 and transfected into an immortalized XP-C cell line (XP4PA-SV-EB) that constitutively expresses Epstein-Barr virus nuclear antigen 1. Constitutive expression of this antigen increases the transformation efficiency by Epstein-Barr virus-based vectors tenfold compared with naive cells⁵. Transformed cells were selected by incubation in hygromycin B-containing medium (pEBS7 contains a hygromycin-resistance marker) and subsequently irradiated with ultra-

violet light. This double selection protocol resulted in the isolation of a cell line, designated XPC-H1, that had wild-type sensitivity to ultraviolet light, as demonstrated by cell survival experiments (Fig. 1a). Episomal DNA was recovered from XPC-H1 cells⁶ and amplified by propagation in MC1061 bacterial cells. This plasmid pool was reintroduced into XP4PA-SV-EB cells, and after selection for transformants, ultraviolet survival experiments on this pool of cells, designated XPC-H2, showed that normal ultraviolet resistance had been fully restored (Fig. 1a). Episomal DNA was rescued from XPC-H2 cells, and individual clones were examined by gel electrophoresis for cDNA inserts. Fifty of sixty clones examined contained a 3.5-kilobase (kb) insert, in addition to a 3.3-kb fragment which was subsequently determined to be a duplication of vector sequences. All assayed clones containing the 3.5-kb insert were able fully to complement XP-C cells, as determined by cell survival experiments. To eliminate duplicated vector sequences, the 3.5-kb insert was recloned into pEBS7 and the resulting plasmid designated pXPC-3. This clone could also fully complement XP4PA-SV-EB cells and also complemented the repair deficiency of a second XP-C cell line (XP4RO) to levels that were essentially wild-type, as determined by the restoration of unscheduled DNA synthesis after somatic cell microinjection experiments (Fig. 1b).

To determine the specificity of complementation, pXPC-3 was introduced into cells from XP complementation groups A and F and the Chinese hamster ovary (CHO) repair-deficient mutant UV135, which belongs to ERCC complementation group 5 (ref. 7). Cell survival experiments on transformed cells that had been selected in hygromycin B medium showed that there was no change in ultraviolet sensitivity due to the presence of pXPC-3 (Fig. 2), indicating that the repair complementation effected by this clone was specific to XP-C cells. Results of control experiments showed that a pEBS7 construct containing the chloramphenicol acetyltransferase (*CAT*) gene expressed *CAT* activity in all of the three cell lines in Fig. 2 at a level similar to that in XP4PA-SV.

The nucleotide sequence and the deduced amino-acid sequence of the XP-C complementing clone (XPCC) are shown

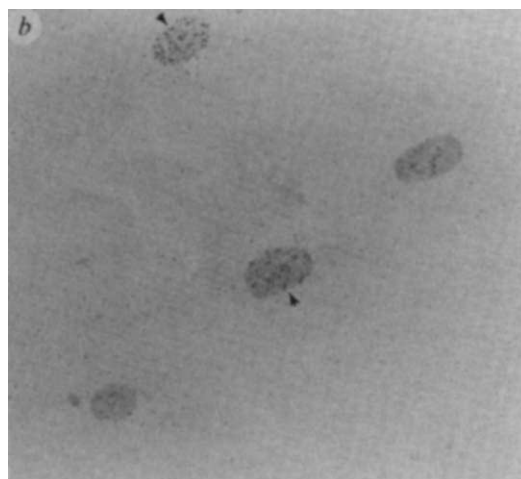
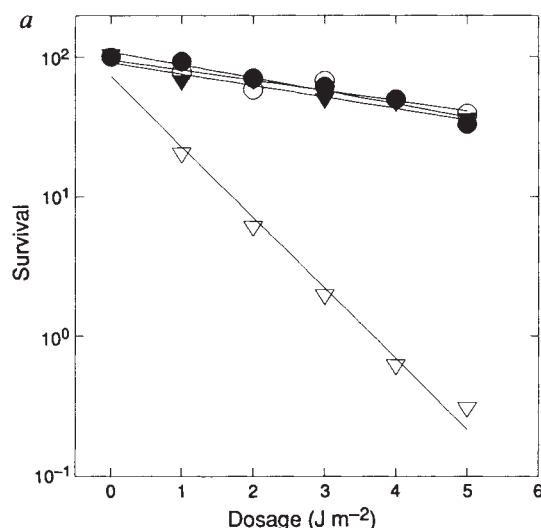


FIG. 1 *a*, Ultraviolet sensitivity of HeLa cells ($\blacktriangledown$), the primary XP-C transfectant XPC-H1 ($\bullet$), the secondary XP-C transfectant XPC-H2 ($\circ$), and XP4PA-SV-EB cells transformed with pEBS7 (∇). *b*, Unscheduled DNA synthesis (UDS) assay following somatic cell microinjection of the group C cell line XP4RO with pXPC-3 DNA. Arrows indicate cells with wild-type levels of UDS after microinjection.

METHODS. *a*, The cell line XP4PA-SV-EB is a derivative of XP4PA-SV (ref. 19) and was isolated as described⁵. A cDNA library containing over 20×10^6 clones was prepared in the vector pEBS7 as described^{5,20}. Twenty 10-cm dishes containing $\sim 5 \times 10^5$ XP4PA-SV cells were transfected with library DNA by calcium phosphate precipitation²¹. After selection for 6 days in medium containing hygromycin B ($75 \mu\text{g ml}^{-1}$), cells were once irradiated with ultraviolet light (2.5 J m^{-2}). After expansion in selective medium, cells were replated at 2×10^6 cells per plate and irradiated with ultraviolet light (4 J m^{-2}) twice at a 24-h interval. After a second expansion, cells were replated at 1×10^6 cells per plate and irradiated with ultraviolet (4 J m^{-2})

three times at 24-h intervals. The surviving cells, which had an almost wild-type level of ultraviolet resistance, were designated cell line XPC-H1. Episomal DNA was recovered from XPC-H1 cells⁶ and electroporated into MC1061 cells²². The pool of recovered plasmids was reintroduced into XP4PA-SV-EB cells, and the transfectants designated cell line XPC-H2. For ultraviolet-survival experiments, a known number of cells were plated into individual dishes, and 24 h later irradiated with light at 254 nm at the indicated dosage. Four days later cells were removed from plates with trypsin, stained with trypan blue, and the number of viable cells determined by counting in a haemocytometer. *b*, XP4RO (group C) cells were plated onto glass coverslips, incubated for 3 weeks in medium containing 0.5% serum to produce quiescent cells, and subsequently injected intranuclearly with pXPC-3 DNA at a concentration of $500 \mu\text{g ml}^{-1}$. After a 16-h incubation, cells were irradiated with light at 254 nm (20 J m^{-2}). Analysis for UDS was as described²³. Wild-type levels of UDS were determined from quiescent WI-38 cells.

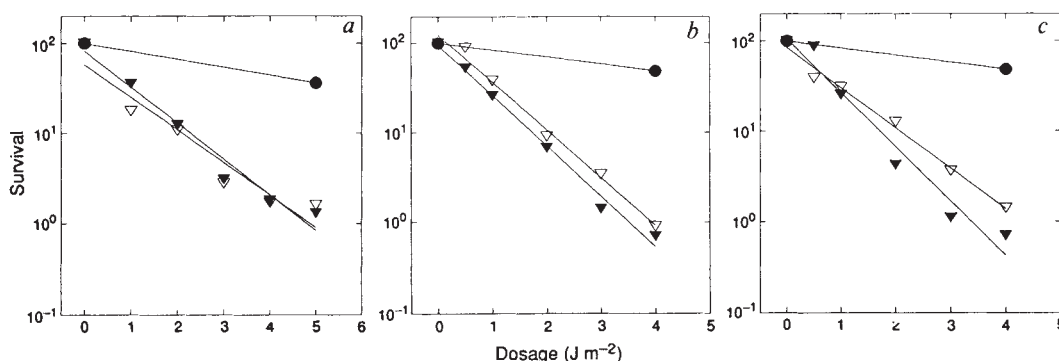
in Fig. 3. The predicted open reading frame of 823 amino acids begins at the fifth ATG in the sequence, located at nucleotide 192. Although it is uncommon for the translation start site to be preceded by four ATG codons, this assignment is likely to be correct because the longest open reading frame initiated by any of the first four ATGs is only 66 base pairs (bp). In addition, according to a matrix scoring method⁸ in which most initiation sites score above +37, only the ATG at position 192 is above this value, having a score of +61. The XPCC sequence has a large, 3' untranslated region of almost 800 bp with a typical polyadenylation site⁹ 16 bp before the beginning of the poly(A) tail and a nearby sequence that is common to short-lived messenger RNAs¹⁰. The XPCC-encoded protein has a predicted M_r of 92,962, is very hydrophilic, having 34.1% charged residues, and is quite basic, with a pI of 8.42. Other significant features include an acidic domain near the amino terminus, two extensive

basic domains, a putative nuclear location signal, and a possible *N*-linked glycosylation site (Fig. 3a). A computer data base search did not identify any nucleic acid or protein species with extensive homology to XPCC, but there was a limited homology with the yeast DNA repair protein RAD4 (refs 11, 12). The overall identity between the two proteins is 23% (44% similarity), but a segment of 230 amino acids (Fig. 3b) near the carboxyl terminus of both proteins has an identity of 30% (52% similarity). This degree of homology indicates that these two genes are distantly related but that RAD4 is not necessarily the yeast homologue of XPCC.

We used northern blot analysis to determine whether expression of the XPCC transcript is altered in XP-C cells (Fig. 4). This revealed a single hybridizing RNA species of ~ 3.8 kb among RNAs from cells with normal repair function and from cells from XP complementation groups A and B. In contrast,

FIG. 2 Ultraviolet sensitivity of repair-deficient cell lines with (∇) or without ($\blacktriangledown$) transformation by the clone pXPC-3. Transfectants were selected by incubation in hygromycin B-containing medium before ultraviolet-survival experiments. *a*, XP12BE-SV (XP group A); *b*, XP3YO-SV (XP group F); *c*, UV-135 (ERCC5). Typical response of normal cells is indicated by the filled circles.

METHODS. Cells were transfected by precipitation with calcium phosphate and transformants were selected by incubation in hygromycin B medium. Ultraviolet-



survival experiments are described in Fig. 1 legend.

METHODS. Fragments of XPCC produced by restriction enzyme digestion or by the exonuclease III method²⁴ were subcloned into the Bluescript II sequencing vector (Stratagene). Sequencing of both strands of XPCC was performed by the dideoxy method on an Applied Biosystems Model 373A automated DNA sequencer.

the 3.8-kb transcript was not detectable or was much reduced among RNAs from four of five XP-C cell lines. A control hybridization with actin as probe indicated that the absence of the XPCC transcript in XP-C cells was not due to degradation of the samples. These and the complementation results suggest that XPCC is the DNA-repair gene defective in XP-C cells. But our results are also consistent with the alternative possibility that the gene we have cloned does not contain the primary defect in XP-C cells, but is rather a factor that is controlled or regulated by the gene in which the actual defect resides.

The XP-C complementation group is the most common form of the syndrome among North Americans and Europeans, accounting for almost 40% of cases⁴. Some progress has been made in defining the biochemical defect in XP-C cells: transcriptionally active regions of the genome seem to be repaired normally, whereas inactive regions, including the non-transcribed strand of active genes, are not repaired to any significant extent^{13,14}. These results suggest that the factor defective in XP-C cells may play a part in damage recognition and/or in altering chromatin structure to allow access by damage-processing

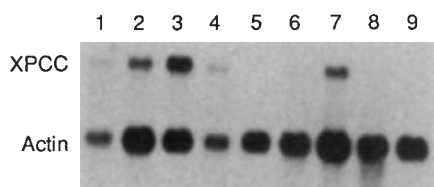


FIG. 4 Northern blot analysis of poly(A)⁺ RNA (5 µg) obtained from the following cell lines: HT-1080 (lane 1), BCG (lane 2), XP11BE-L1 (group B) (lane 3), XP12BE-SV (group A) (lane 4), XP4PA-SV (group C) (lane 5), XP1BE-L1 (group C) (lane 6), XP8BE-L1 (group C) (lane 7), XP1MI (group C) (lane 8), and XP3BE-L3 (group C) (lane 9).

METHODS. Poly(A)⁺ RNA was extracted from cells with a Fast Track kit (Invitrogen). Preparation of membranes and hybridizations were carried out essentially as described²⁵. Probe was prepared by digestion of pXPC-3 with *Sfi*I and gel purification of the single 3.5-kb band. DNA was labelled by the random priming method.

enzymes. The sequence of the XPCC protein reported here provides few new clues as to its biochemical function, although the presence of both highly acidic and basic domains permits interaction with chromosomal proteins and DNA. Isolation of the gene should facilitate the study of its function in DNA excision repair and its involvement in mutagenesis and carcinogenesis in human cells.

Isolation of DNA repair genes by direct complementation of XP cells through genomic DNA transfection procedures has proved difficult and so far only the cloning of the XPAC gene has been achieved¹⁵, although genes that complement cells from XP groups B and D have been isolated through the use of repair-deficient rodent cell lines^{16–18}. Our expression cloning approach has circumvented these problems, being based on an Epstein-Barr virus episomal vector that transforms human cells with high efficiency and is easily rescued for amplification in bacterial cells. The efficacy of this system is demonstrated by the cloning of the XPCC gene described here, and it should be generally applicable for isolating genes by direct phenotypic selection in human mutant cell lines. □

The Pas2 protein essential for peroxisome biogenesis is related to ubiquitin-conjugating enzymes

Franziska F. Wiebel & Wolf-H. Kunau

Institut für Physiologische Chemie der Ruhr-Universität Bochum, Abteilung für Zellbiochemie, D-4630 Bochum 1, Germany

IN the yeast *Saccharomyces cerevisiae*, PAS genes are essential for the biogenesis and proliferation of peroxisomes^{1,2}. Recently, the first two genes, *PAS1* (ref. 3) and *PAS3* (ref. 4), have been characterized. Here we report the cloning and sequencing of the *PAS2* gene. It encodes a new member of the ubiquitin-conjugating (UBC) protein family^{5–7} and is the first member associated with peroxisomes. The proposed function of the *Pas2* protein as a UBC enzyme (UBC10) is supported by the fact that site-directed mutagenesis of a strictly conserved and functionally essential cysteine residue of UBC proteins leads to mutant *Pas2* proteins unable to complement *pas2* mutant strains. Ubiquitination of proteins is known to play an important part in DNA repair, sporulation, cell cycle control and degradation of abnormal proteins^{5–8}. We provide evidence for a crucial role of the ubiquitin-conjugation pathway in organelle formation.

Of the 18 known complementation groups⁹ of *pas*-mutants in *S. cerevisiae*, most belong to the type I subclass² characterized by: (1) the absence of functional peroxisomes, resulting in (2) an inability to grow on oleic acid as a sole carbon source, and (3) mislocalization of peroxisomal matrix enzymes which are found in an active form in the cytosol. Defects leading to type I *pas*-phenotype are usually expected to be due to impairments in the peroxisomal import machinery². The original *pas2* mutant¹ and the *pas2* null mutant exhibit type I phenotype.

The *PAS2* gene (Fig. 1a) encodes a polypeptide of 183 amino acids (*M_r* 21.118K). A shorter fragment, starting at nucleic-acid position 37 of the open reading frame, encodes a protein with 165 residues (18.902K) which still complements the *pas2*-mutation when expressed from a *CycGal*-promoter. The corresponding ATG is preceded by another in a different open reading frame (Fig. 1)¹⁰.

A search of databases revealed that the *PAS2* protein (*Pas2*) has strong sequence similarity (45–60%, Fig. 1b) to the UBC proteins^{5–7} in a highly conserved segment of about 140 residues including the essential 'active site' cysteine residue (Fig. 1b) responsible for thioester formation with ubiquitin¹¹. The alignment suggests the second methionine codon of the *PAS2* open reading frame as the physiological start of translation. To see whether the presumed 'active site' cysteine residue is indeed indispensable for *Pas2* protein function, its codon was mutated to serine and alanine, respectively. The corresponding gene products, although expressed from a strong *CycGal*-promoter and immunologically detectable in the cell extracts (Fig. 2), were unable to complement the null mutant of *PAS2* (Table 1). This supports the conclusion that *Pas2* protein is indeed a UBC protein.

UBC proteins are part of the ubiquitin-conjugation cascade^{5–8,12} which transfers ubiquitin to protein substrates by means of a general ubiquitin-activating enzyme (UBA/E1) and different ubiquitin-conjugating enzymes (UBC/E2), many of which cooperate with ubiquitin-protein ligases (UBR/E3)⁶. One well known function of the ubiquitin pathway is to mark substrates for subsequent degradation^{12,13}. A direct involvement in protein degradation has been shown for UBC1, UBC4 and UBC5 of *S. cerevisiae*^{13–15}. Others relate to more complex cellular processes such as DNA repair, induced mutagenesis and sporulation (UBC2), cell cycle control (UBC3) or the secretory pathway (UBC6)⁷. In all of these cases, it is not yet clear how the specific cellular function is related to the ubiquitination signal.

Received 18 March; accepted 7 July 1992.

1. Cleaver, J. E. & Bootsma, D. A. *Rev. Genet.* **26**, 365–369 (1975).
2. Andrews, A. D., Barrett, S. F. & Robbins, J. H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1984–1988 (1978).
3. Cleaver, J. E. *Nature* **218**, 652–655 (1968).
4. Cleaver, J. E. & Kraemer, K. H. in *The Metabolic Basis of Inherited Disease* 6th edn (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 2949–2971 (McGraw-Hill, New York, 1989).
5. Peterson, C. & Legerski, R. *Gene* **107**, 279–284 (1991).
6. Hirt, B. *J. molec. Biol.* **26**, 365–369 (1967).
7. Thompson, L. H., Busch, B. D., Brookman, K., Mooney, C. L. & Glaser, D. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3734–3737 (1981).
8. Stormo, G. D. in *Nucleic Acid and Protein Sequence Analysis: A Practical Approach* (eds Bishop, M. J. & Rawlings, C. J.) 231–258 (IRL, Oxford, 1987).
9. Wickens, M. & Stephenson, P. *Science* **226**, 1045–1051 (1984).
10. Shaw, G. & Kamen, R. *Cell* **46**, 659–667 (1986).
11. Gietz, R. D. & Prakash, S. *Gene* **74**, 535–541 (1988).
12. Couto, L. B. & Friedberg, E. C. *J. Bact.* **171**, 1862–1869 (1989).
13. Kantor, G. J., Barsalou, L. S. & Hanawalt, P. C. *Mut. Res.* **243**, 171–180 (1990).
14. Venema, J. et al. *Molec. cell. Biol.* **11**, 4128–4134 (1991).
15. Tanaka, K., Satokata, I., Ogita, Z., Uchida, T. & Okada, Y. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5512–5516 (1989).
16. Weeda, G. et al. *Cell* **62**, 777–791 (1990).
17. Weber, C. A., Salazar, E. P., Stewart, S. A. & Thompson, L. H. *Molec. cell. Biol.* **8**, 1137–1146 (1988).
18. Flejter, W. L., McDaniel, L. D., Johns, D., Friedberg, E. C. & Schultz, R. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 261–265 (1992).
19. Daya-Grosjean, L., James, M. R., Drougard, C. & Sarasin, A. *Mut. Res.* **198**, 185–196 (1987).
20. Aruffo, A. & Seed, B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8573–8577 (1987).
21. Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745–2752 (1987).
22. Dower, W. J., Miller, J. F. & Ragsdale, C. W. *Nucleic Acids Res.* **16**, 6127–6135 (1988).
23. Legerski, R. J., Brown, D. B., Peterson, C. A. & Roberson, D. L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5676–5679 (1984).
24. Henikoff, S. *Meth. Enzym.* **155**, 156–165 (1987).
25. Virca, G. D., Northemann, W., Shiels, B. R., Widera, G. & Broome, S. *Biotechniques* **8**, 370–371 (1990).

ACKNOWLEDGEMENTS. We thank R. Meyn and M. Weil for critiques of the manuscript. This work was supported in part by grants from the US National Cancer Institute and the M.D. Anderson Cancer Center Development Fund.

METHODS. The *PAS2* gene was isolated from a *S. cerevisiae* genomic bank contained in the vector Ycp50 by functional complementation of the *pas2* mutant. Complemented strains regained wild-type phenotype with regard to all three criteria (see text, Table 1). Authenticity of the isolated clone was confirmed by crossing the gene deletion mutant with the original mutant, to give diploids and subsequently complete tetrads displaying *pas* phenotype. The sequence of both strands was determined using a T7 sequencing kit (Pharmacia) and vector pBLUESCRIPT (Stratagene). For site-directed mutagenesis, a 385-bp fragment was generated by polymerase chain reaction using two oligonucleotides, one of which covered the Cys 115 codon and was degenerated to yield Ser 115 and Ala 115. An internal 360-bp *Bst*XI-fragment was then used to exchange the wild-type sequence of existing *PAS2*-constructs.

Subcellular fractionation studies (Fig. 3) showed that when transformants display a wild-type phenotype (Table 1), the fusion constructs migrate at a density of 1.22 g cm^{-3} together with catalase, suggesting these are associated with peroxisomes. This is the case for the complemented *pas2* null mutant (Fig. 3a) as well as wild-type cells transformed with either construct

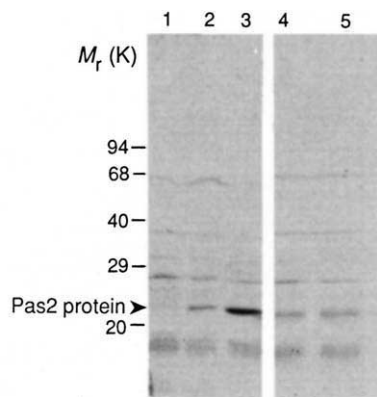


FIG. 2 Immunological detection of Pas2 protein (arrow). Molecular markers are indicated on the left. Pas2 protein is not detectable in wild-type yeast (UTL-7A), or in single-copy (YCp50) or multi-copy (YE352, lane 1) transformants of the *pas2* mutant. But Pas2 protein can be detected in whole-cell lysates when the protein (183-residue open reading frame) is expressed from single-copy (pBM272, lane 2) or multicopy (YE352, containing the promoter fragment of pBM272, lane 3) plasmids under control of the strong GAL1-promoter under inducing conditions. Lanes 4 and 5 show detection of the gene product of the mutagenized *PAS2* yielding Pas2^{Ala} and Pas2^{Ser}, respectively, expressed from vector pEMBLyex4 in the mutant strain. METHODS. Antibody production, preparation of samples and immunoblotting was done as described in ref. 4. For preparation of samples, cells were grown for 18 h on oleic acid medium (Table 1).

(data not shown). But, in gradients of noncomplemented *pas*-mutant strains (Table 1; *pas2* containing *PAS2*^{Ser}-*lacZ*, Fig. 3b, or transformed *pas5*, data not shown), peak activity of β -galactosidase is found at a much lower density of 1.14 g cm^{-3} whereas peroxisomal matrix enzymes like catalase are found in the upper gradient fractions (Fig. 3b). For both phenotypes, the distribution pattern of β -galactosidase is matched by that of Pas3 protein, a peroxisomal integral membrane protein⁴ (Fig. 3b).

The results indicate that the association of the fusion protein with peroxisomal fractions is independent of Cys 115, and therefore of Pas2 protein activity. The observation that β -galactosidase activity is found at different densities depending on the phenotype excludes the possibility of incidental migration of fusion protein aggregates at a density typical of intact peroxisomes. Moreover, the comigration with Pas3 protein suggests that the fusion proteins are indeed associated with peroxisomes as well as peroxisomal structures in the mutant phenotype, which might in several respects be comparable to the so-called ghosts of Zellweger fibroblasts¹⁶.

It can be concluded that the Pas2 protein portion of either fusion protein is able to direct it to intact peroxisomes and presumably peroxisomal membrane structures, thus serving as a peroxisomal membrane marker. That *PAS2*^{Cys}-*lacZ* confers complementation, supports the notion that the fusion protein is directed to the true intracellular locus of Pas2 protein, thus indicating that Pas2 protein is a peroxisomal protein.

Pas2p/UBC10 provides the first example of a *UBC* gene product required for the formation of a subcellular compartment, the peroxisome. Its proposed localization suggests that the crucial ubiquitination step takes place at the organelle. Recently, a ubiquitin requirement for the insertion of monoamine oxidase A into the mitochondrial outer membrane has been demonstrated using *in vitro* assays¹⁷, and studies on E1-dependent lysosomal protein degradation and on neurodegenerative diseases led to proposals that the ubiquitin pathway is involved in vacuolar biogenesis or protein uptake^{8,18,19}. Studies on the *PAS2* gene and its product might therefore prove to be of more general significance, concerning the linkage of the ubiquitin pathway to eukaryotic compartmentation.

A possible role of ubiquitination by Pas2 protein could be to mediate the degradative elimination of a factor inhibiting peroxisome biogenesis². Cyclin and p53 are examples of regulatory proteins degraded in the ubiquitin pathway⁸. Alternatively, the ubiquitin tag could act by regulating the activity of specific import factors or by converting protein substrates into import-competent forms.

We should now look for potential ubiquitination substrates of Pas2/UBC10 to help elucidate not only the mechanisms of peroxisome biogenesis but also an example of the involvement of a ubiquitination event in a complex cellular function. Pas2 protein is a promising candidate because it is interrelated with the growing number of described peroxisomal mutants of

TABLE 1 Distribution pattern of marker enzymes

Strain	Enzyme	Activity* in 25,000g		A1/A2
		Supernatant fraction (A1)	Pellet fraction (A2)	
<i>PAS2::LEU2/YCp50</i>	Catalase	479.2	3.0	160
	Thiolase	83.4	14.1	5.9
	Epimerase	89.7	14.0	6.4
	Fumarase	54.3	68.8	0.8
<i>PAS2::LEU2/YCp50-PAS2</i>	Catalase	10.0	47.5	0.2
	Thiolase	8.5	17.7	0.5
	Epimerase	5.8	20.2	0.3
	Fumarase	30.8	37.5	0.8
<i>PAS2::LEU2/pEMBLyex4-PAS2^{Ser}</i>	Catalase	479.8	4.9	98
	Thiolase	122.5	9.5	12.9
	Epimerase	243.7	21.3	11.4
	Fumarase	7.5	16.7	0.5
<i>PAS2::LEU2/PAS2^{Cys}-lacZ</i>	Catalase	190.9	66.4	2.9
	Thiolase	42.0	44.0	0.9
	Epimerase	76.8	53.8	1.4
	Fumarase	21.7	51.7	0.4
	β -Galactosidase	170.0	1,498.3	0.1
Wild-type (UTL-7A)	Catalase	10.3	33.8	0.3
	Thiolase	7.7	10.9	0.7
	Epimerase	6.1	31.7	0.2
	Fumarase	15.8	29.8	0.5
	β -Galactosidase	12.3	11.7	1.1
Wild-type (UTL-7A)/ <i>PAS2^{Ser}-lacZ</i>	Catalase	26.6	43.1	0.6
	Thiolase	6.8	10.6	0.6
	Epimerase	9.1	32.0	0.3
	Fumarase	9.2	20.8	0.4
	β -Galactosidase	58.3	815.7	0.1
<i>PAS2::LEU2/PAS2^{Ser}-lacZ</i>	Catalase	371.5	1.9	192
	Thiolase	40.2	3.0	13.4
	Epimerase	79.3	11.3	7.0
	Fumarase	16.3	39.8	0.4
	β -Galactosidase	451.7	1650	0.3
<i>PAS5::LEU2/PAS2^{Cys}-lacZ</i>	Catalase	356.8	1.7	212
	Thiolase	15.4	2.1	7.3
	Epimerase	53.3	7.3	7.3
	Fumarase	9.8	45.3	0.2
	β -Galactosidase	218.3	558.3	0.4

Distribution pattern of peroxisomal and mitochondrial marker enzymes in the 25,000g supernatant and pellet fractions of cell lysates from different strains. Cells were grown for 15 h on oleic acid media (YNO (ref. 1) containing 0.1% glucose, 0.1% yeast extract) inoculated from SD-precultures containing 0.3% glucose to $1-2 \times 10^6$ cells per ml. Preparation of yeast lysates and fractionation by differential centrifugation was done as described in ref. 1. Fumarase activity distribution serves as a control for the intactness of organelles. For enzyme assays see refs 4 and 20. *PAS2*^{Cys}-*lacZ* and *PAS2*^{Ser}-*lacZ* are constructs in vector YE358R. Values obtained for *PAS2*^{Ser}-*lacZ* overexpressed in the wild-type strain demonstrates that the mutation is not dominant.

* Catalase (μkat), β -galactosidase (pkat), others (nkcat).

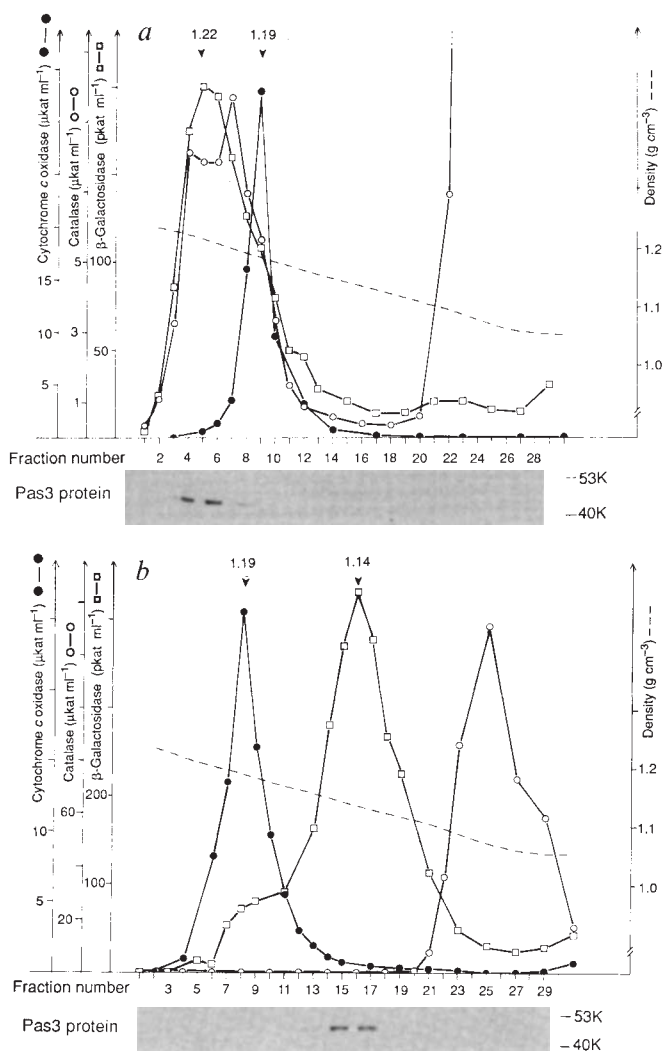


FIG. 3 Activities of catalase, cytochrome *c* oxidase, and β -galactosidase, and immunological detection of Pas3 protein in fractions of a continuous 20–53% sucrose density gradient of cell lysates of the *pas2* null mutant transformed with *PAS2^{Cys}-lacZ* (a) and *PAS2^{Ser}-lacZ* (b) (see text). Densities at activity peaks of cytochrome *c* oxidase and β -galactosidase are indicated. Virtually the same distribution as in a was obtained for wild-type (UTL-7A) transformed with either *PAS2^{Cys}-lacZ* or *PAS2^{Ser}-lacZ*, and as in b for the *pas5* null mutant transformed with *PAS2^{Cys}-lacZ*.

METHODS. For the growth of cells, preparation of the 25,000g organellar pellet and enzyme assays see Table 1. Sucrose gradient fractionation was described in ref. 4.

increasingly well defined phenotype, some of which may provide the immediate molecular context of Pas2 protein. □

Received 16 April; accepted 17 July 1992.

- Erdmann, R., Veenhuis, M., Mertens, D. & Kunau, W.-H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5419–5423 (1989).
- Höfheld, J., Mertens, D., Wiebel, F. F. & Kunau, W.-H. in *New Comprehensive Biochemistry Series: Membrane Biogenesis and Protein Targeting* (eds Neupert, W. & Lill, R.) (Elsevier, New York, in the press).
- Erdmann, R. *et al. Cell* **64**, 499–510 (1991).
- Höfheld, J., Veenhuis, M. & Kunau, W.-H. *J. Cell Biol.* **114**, 1167–1178 (1991).
- Jentsch, S., Seufert, W., Sommer, T. & Reins, H.-A. *Trends biochem. Sci.* **15**, 195–198 (1990).
- Jentsch, S., Seufert, W. & Hauser, H.-P. *Biochim. biophys. Acta* **1089**, 127–139 (1991).
- Jentsch, S. *Trends Cell Biol.* **2**, 98–103 (1992).
- Finley, D. & Chau, V. *A. Rev. Cell Biol.* **7**, 25–69 (1991).
- Erdmann, R. & Kunau, W.-H. *Cell Biochemistry and Function* (eds Crook, E. M. & Chayen, J.) (Wiley, Chichester, in the press).
- Deshaies, R. J. & Schekman, R. *J. Cell Biol.* **109**, 2653–2664 (1989).
- Sung, P., Prakash, S. & Prakash, L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2695–2699 (1990).
- Hershko, A. *Trends biochem. Sci.* **16**, 265–268 (1991).
- Johnson, E. S., Bartel, B., Seufert, W. & Varshavsky, A. *EMBO J.* **11**, 497–505 (1992).
- Seufert, W. & Jentsch, S. *EMBO J.* **9**, 543–550 (1990).

- Seufert, W., McGrath, J. P. & Jentsch, S. *EMBO J.* **9**, 4535–4541 (1990).
- Santos, M. J., Imanaka, T., Shio, H. & Lazarow, P. B. *J. biol. Chem.* **263**, 10502–10509 (1988).
- Zhuang, Z., Marks, B. & McCauley, R. B. *J. biol. Chem.* **267**, 591–596 (1992).
- Gropper, R. *et al. J. biol. Chem.* **266**, 3602–3610 (1991).
- Laszlo, L., Doherty, F. J., Osborn, N. U. & Mayer, R. J. *FEBS Lett.* **261**, 365–368 (1990).
- Rose, M. D., Winston, F. & Hieter, P. *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press, New York, 1990).
- Reynolds, P., Weber, S. & Prakash, L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 168–172 (1985).
- Goel, M. G. *et al. Science* **241**, 1331–1335 (1988).
- Qin, S., Nakajima, B., Nomura, M. & Arfin, S. M. *J. biol. Chem.* **266**, 15549–15554 (1991).

ACKNOWLEDGEMENTS. We thank J. Höfheld for helpful comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft.

Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor

Colin L. Stewart*, Petr Kaspar†, Lisa J. Brunet*, Harshida Bhatt*, Inder Gadi‡, Frank Köntgen§ & Susan J. Abbonando*

* Department of Cell and Developmental Biology, Roche Institute of Molecular Biology, Roche Research Center, 340 Kingsland Street, Nutley, New Jersey 07110, USA

† Institute of Molecular Genetics, Flemingovo 2, Prague 6, Czechoslovakia

‡ Roche Biomedical Laboratories Inc., 1 Roche Drive, Raritan, New Jersey, USA

§ F. Hoffmann-La Roche Ltd, Pharmaceutical Research, Basle New Technologies, CH4002 Basle, Switzerland

A CRITICAL point during mammalian pregnancy is the implantation of the blastocyst when the embryo attaches to the wall of the uterus. The autonomously developing preimplantation embryo then becomes dependent on the maternal environment for its continued development. Little is known about the regulation of implantation, except that a complex interaction between peptide and steroid hormones synchronizes the preparation of the uterus for implantation with the development of the embryo. Whether the implantation event is under maternal or embryonic control is also unclear (reviewed in refs 1, 2). We have previously shown that a cytokine, leukaemia inhibitory factor (LIF), is expressed in the uterine endometrial glands specifically on the fourth day of pregnancy³. This burst of expression is under maternal control and always precedes implantation of the blastocyst. Here we report that transient expression of LIF in mice is essential for implantation. Females lacking a functional LIF gene are fertile, but their blastocysts fail to implant and do not develop. The blastocysts, however, are viable and, when transferred to wild-type pseudopregnant recipients, they can implant and develop to term.

The gene encoding LIF was mutated by introducing a phosphoglycerate kinase (PGK)neo poly(A)⁺ cassette into a *Sma*I site in the third exon. This disrupts the reading frame of the gene, resulting in a truncated LIF protein lacking the carboxy-terminal 81 amino acids, including the last nine that are essential for its biological activity⁴. This construct (Fig. 1a) was used to mutate the endogenous LIF gene in embryonic stem (ES) cells by targeted mutagenesis using homologous recombination^{5–8}. ES clones in which the mutant gene had homologously recombined with the endogenous LIF gene were initially identified by polymerase chain reaction (PCR) screening and confirmed by Southern blotting (Figs 1 and 2). Six homologous recombinant clones were isolated, one of which was then used to produce ES chimaeras by injection into recipient blastocysts. Heterozygous offspring were derived from the founders and were used to produce individuals homozygous for the mutated LIF allele (Fig. 2). Viable adult homozygotes were obtained (Table 1a) that consistently showed a retarded postnatal growth rate which resulted in their being 25–35% smaller (in body weight) than their wild-type or heterozygous siblings (data not shown).

To confirm that the mutated allele disrupted production of protein that was biologically active, primary mouse embryo fibroblast (PMEF) cultures were established from wild-type, heterozygous and homozygous 13-day embryos (day 1 is day of plug). These were used to determine whether they could support ES cell growth because ES cells are dependent on LIF produced by PMEFs^{9,10}. Fibroblasts homozygous for the mutant LIF allele were unable to sustain the growth of undifferentiated ES cells and no colonies were formed. Heterozygous lines of PMEFs supported ES cell colony formation to a lesser extent, with about one-third the number of undifferentiated clones obtained compared to the wild-type lines (Fig. 3). Presumably, this was because these fibroblasts, which had only one functional LIF gene, produced less than 1,000–2,000 IU ml⁻¹, the minimum amount required to prevent ES cell differentiation⁹. Furthermore, northern analysis of the LIF RNA transcripts showed a reduction in signal intensity of the 4.2-kilobase (kb) transcript

in the heterozygotes and an absence of intact LIF transcripts from the homozygous PMEFs (Fig. 3).

LIF regulates the growth and differentiation of a variety of cell types including those of embryonic origin, such as ES cells^{9,10} and primordial germ cells¹¹. We therefore investigated, in the homozygotes, the effect of LIF deficiency on the reproduction of these mice. Males homozygous for LIF deficiency were fertile and able to sire offspring from both wild-type and heterozygous females (Table 1a). Females, homozygous for LIF deficiency, could also mate with wild-type males of proven fertility, but none of the six mice tested became pregnant despite repeated matings. But subsequent analysis of female homozygotes, mated with either heterozygous or homozygous males, revealed that they were fertile. Morphologically normal blastocysts from matings to either type of male were found in the uterine lumen on the fourth day of gestation. Blastocysts recovered from the homozygous × homozygous matings, were transferred to 3-day pseudopregnant wild-type recipients, where they implanted and developed to term (Table 1b). Analysis of female homozygotes on day 7 of gestation revealed that blastocysts were still present in the uterine lumen. The blastocysts were not surrounded by a zona pellucida, and there was no overt indication that implantation had occurred, or any evidence of a decidual response in the uterus which should have started on the fifth day of gestation. The uteri from these females were pale and poorly vascularized, unlike those of normal seventh day pregnancies. We are now doing a detailed histological analysis of the uteri in homozygotes. Some of these day-7 blastocysts derived from matings with homozygous males were transferred to 3-day pseudopregnant wild-type recipients, where they implanted and developed to term (Table 1b).

In the complete absence of LIF of either maternal or embryonic origin, mouse embryos are capable of developing to the blastocyst stage with the formation of a trophoderm and inner cell mass. But uterine expression of LIF from the endometrial glands on the fourth day of gestation is essential for implantation. In its absence, the blastocysts are viable for at least 2–3

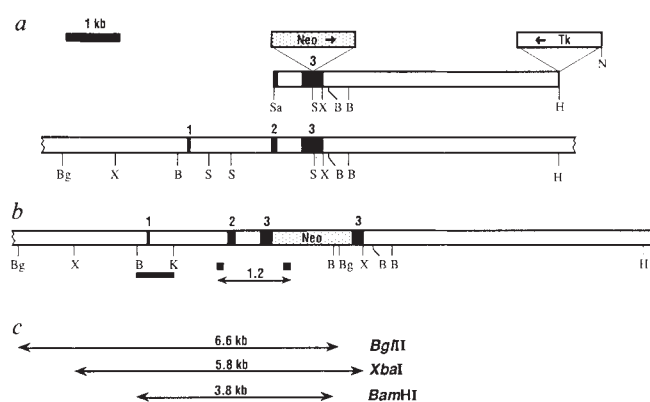


FIG. 1 Procedure used to mutate the endogenous LIF gene by homologous recombination using positive-negative selection²⁵ with a LIF vector. A partial copy of the LIF genomic gene in which the 5' end of the gene started at a *SalI* site just before the second exon and extended 5.6 kb to a *HindIII* site was isolated from a λ D-*Sau3A* genomic library from a CD1 mouse (provided by R. Kinloch) with a murine LIF complementary DNA probe. Confirmation that the LIF gene had been isolated was obtained by comparison with the published sequence²⁶. a, the *SalI*-*XbaI* (*Sa*-*X*) fragment was subcloned, and the PGKneo poly(A)⁺ plasmid (from M. McBurney) was blunt-end ligated, in the same orientation as the LIF gene into the *SmaI*(S) site in the third exon. The *SalI*-*XbaI* PGKneo-containing fragment was ligated to the *XbaI*-*HindIII* (X-H) segment of the LIF gene that encoded the entire 3'-untranslated sequences that had been cloned into a pKS plasmid containing a herpes simplex thymidine kinase (TK) cassette 3' to the *HindIII* site. The vector was linearized (N) by *NotI* digestion, and 3×10^7 W.9.3 ES cells (derived from a 129/Sv blastocyst by J. Mann) were electroporated at 400 V/250 μ F using a BioRad gene pulser. The electroporated ES cells were plated onto primary embryo fibroblasts, derived from a transgenic mouse line that constitutively expresses the neomycin resistance gene²⁷. G418 selection using 350 μ g total geneticin ml⁻¹ (Gibco) was started 2 days after plating, and gancyclovir selection at 2×10^{-7} M at 5 days after plating. Gancyclovir resulted in selection against G418 clones retaining the TK cassette²⁵. The surviving clones were individually picked 11 days after plating, and expanded in 48-well microtitre dishes. Clones were screened 3 days later for homologous recombinants by PCR with the following primers (small rectangular symbols in b): 5' primer, 5'-GGATTGTGCCCTTACTGCTGC-3'; and the 3' primer in the PGK promoter, 5'-GCACGAGACTAGTGAGACGTG-3', which amplified a 1.2-kb fragment in the homologous recombinants. The PCR reaction was done in the following buffer: 25 mM KCl, 67 mM Tris-HCl, pH 8.8, 1 mM MgCl₂, 0.2% Nonidet P-40, 0.2% Tween-20, 0.1 mg ml⁻¹ BSA, with the following times and temperatures: 95 °C for 20 s, 65 °C for 30 s and 72 °C for 30 s (a total of 40 cycles)²⁸. From a total of 96 clones screened, six were homologous recombinants. This was confirmed by restriction enzyme digestion and Southern blotting³¹ using the *Bam*HI-*Kpn*I fragment containing the first exon as a probe (b), with subsequent confirmation using an *XbaI*-*Sma*I probe from the third exon to ensure that tandem integration had not occurred and that the endogenous wild-type gene had been replaced by its mutated homologue.

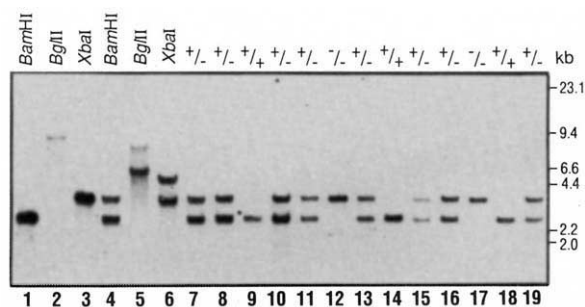


FIG. 2 Six clones from the W9.3 line carried a mutated LIF allele as confirmed by Southern blotting. But when analysed, all had lost the Y chromosome and thus had a karyotype of 39XO chromosomes. One of these clones, W9.23.17.3, was injected into 26 C57BL/6J blastocysts which were transferred to pseudopregnant recipients²⁹. Sixteen offspring were born with 14 surviving to term, of which 11 (8♀ and 3♂) were chimaeras, revealed by their agouti coat colour. All eight females were bred with C57BL/6J males and five produced heterozygous offspring. Lanes 1–3, *Bam*HI, *Bgl*II and *Xba*I digests of W9.3 DNA showing the expected band sizes of the recombinant bands detected by the *Bam*HI-*Kpn*I probe spanning the first exon (probe shown in Fig. 1b) and sizes of the recombinant bands in Fig. 1c). Lanes 4–6, same digests from the W9.3.17.3 homologous recombinant showing the 3.8 *Bam*HI, 6.6 *Bgl*II and 4.5 kb *Xba*I recombinant LIF bands. Lanes 7–10, 4 agouti (therefore ES-derived) offspring from one of the female chimaeras. The three individuals in lanes 7, 8 and 10 contain the recombinant LIF allele shown by the presence of the 3.8-kb *Bam*HI band. Lanes 11–20, 3-week-old offspring derived from intercrossing two heterozygotes, with individuals in lanes 12 and 17, showing that they are homozygous for the mutant LIF allele because they lack the wild-type 3.0-kb and only contain the 3.8-kb recombinant band.

days beyond the time they should have implanted. It is not yet known how LIF acts. It may prepare the endometrial tissue for blastocyst implantation and/or it may act on the mouse blastocyst expressing LIF receptors (D. Hilton, personal communication).

A complication in determining how LIF functions is that blastocysts may synthesize their own LIF^{12,13}. Thus, females deficient in LIF could possibly, because of sufficient LIF being produced by the blastocysts, still support the implantation of wild-type or even heterozygous blastocysts. But this is unlikely, because homozygous females mated to wild-type males never became pregnant and wild-type blastocysts transferred to pseudopregnant homozygotes never implanted (Table 1c).

An indication that LIF might at least be acting on the uterus was suggested by the inability of wild-type blastocysts, incubated *in vitro* in the presence of 2×10^3 IU ml⁻¹ LIF for 24–48 h before transfer, to implant into the uteri of homozygous pseudopregnant recipient females. Thus, it is unlikely that LIF is directly affecting the blastocyst by inducing it to implant. Transfer to wild-type recipients did result in implantation (Table 1c). In addition, a single injection into the uterine lumen of between 10 and 20 μ l ($1\text{--}2 \times 10^5$ IU) of recombinant murine LIF at the

time of transfer, also failed to stimulate implantation in the homozygotes (Table 1d). Injected LIF, however, has a very high clearance rate from the circulatory system ($t_{1/2}$ of 3–5 min)¹⁴ and a single injection may not provide a sufficient concentration for long enough to induce implantation.

In an attempt to circumvent this, a micro-osmotic pump (Alza Corp., Palo Alto, California) that delivers recombinant LIF at a rate of 10^4 IU per hour for about 72 h was implanted into the peritoneum of 3-day pregnant homozygous females mated to homozygous males. A total of seven females received the pumps. Examination on the sixth day of gestation revealed that three mice had observable implantation sites (a combined total of 11 were detected). The remaining four recipients had no implantation sites, but were not pregnant because no blastocysts were recovered from their uteri. The uteri from two of the females having implantation sites were examined histologically. This revealed that in all six sites a decidual response was occurring and two of the more advanced sites contained embryos. The third female (having five implantation sites) was reexamined on the tenth day of gestation, but all the implantation sites had disappeared. Three additional females, as controls, also received pumps containing phosphate buffered saline and no LIF. No implantation sites were detected on the sixth day of gestation, but 14 blastocysts were recovered from their uteri. These results showed that the availability of LIF for the duration over which implantation occurs, can result in a decidual response occurring with embryos undergoing implantation. It is, however, not known if continuous expression of LIF, which occurs at low levels throughout pregnancy³, is required to maintain embryonic development.

A requirement for LIF, at least *in vitro*, has been demonstrated in a variety of other biological systems such as neuronal differentiation/maintenance, osteoblast formation and haemopoiesis (reviewed in ref. 15). That viable homozygotes deficient in LIF expression have been derived suggests that at least for the first few months of postnatal existence there is no absolute requirement for LIF *in vivo*. But its absence could lead to some abnormal phenotype or some subtle alteration which may not manifest itself until later; this is being investigated.

LIF exhibits some homology with other secreted factors¹⁶, in particular oncostatin M (ref. 17) and ciliary neurotrophic factor (CNTF), both of which may bind to the same receptor

TABLE 1 Reproductive frequencies in LIF mutant mice

(a)		Numbers born			
♀	♂	born	+/+	+/-	-/-
+/- (3)	× +/-	27	7 (26%)	14 (52%)	6 (22%)
+/- (4)	× -/-	35	0	18 (51%)	17 (49%)
-/- (6)	× +/+	0	—	—	—

(b)		Numbers born			
Female homozygotes ♀	♂	Number of blastocysts isolated	Day of gestation	Numbers born following transfer to +/- D3 recipients	
-/- (1)	× +/+	9	4	8	
-/- (1)	× -/-	24 (SO)	4	7	
-/- (2)	× +/-	11	7	2	
-/- (5)	× -/-	31	7	9	

(c) Transfer of wild-type blastocysts incubated for 24–48 h in 10^4 IU LIF		
Number transferred	Recipient (numbers)	Number of implantation sites, day 7
52	+/+ (5)	18
53	-/- (5)	0

(d) Transfer to wild-type blastocysts simultaneously with 10^5 IU LIF		
Number transferred	Recipient (numbers)	Numbers of implantation sites, day 7
27	+/+ (6)	13
31	-/- (6)	0

a, Reproductive rates of LIF-dependent heterozygotes and homozygotes. All offspring were screened by PCR analysis of DNA from toe biopsies. The results were confirmed by Southern blotting. The figure in parentheses is the number of individual female homozygotes that were test mated, each being mated at least twice with wild-type males of proven fertility. b, Female homozygotes are fertile when mated with either wild-type or homozygous males. Blastocysts were flushed from the uteri on day 4 of gestation (day 1 is day of plug) and transferred to day-3 wild-type pseudopregnant recipients. SO, This female was 1 of 2 that was superovulated by an injection of pregnant mare serum and human chorionic gonadotrophin before being mated to a homozygous male. The other female did not mate. The figures in parentheses are the number of females used to obtain the blastocysts. All individuals born after blastocyst transfer were screened by PCR to confirm their expected genetic status. c, Wild-type blastocysts from (C57BL/6J × CBA/J)_{F₁} × C57BL/6J matings were isolated on day 3 of gestation (8-cell stage) and incubated overnight, either with or without 10^4 IU of murine recombinant LIF (ESGRO, Gibco-BRL) for either 24 or 48 h (day-4 or day-5 blastocysts, respectively). The blastocysts were transferred to the uteri of pseudopregnant wild-type recipients or pseudopregnant LIF-deficient homozygotes which were either at day 3 or 4 of gestation. After transfer, the recipient was then analysed for the presence of implantation sites on day 7, equivalent to the seventh day of gestation. d, Wild-type blastocysts (day 4) from (C57BL/6J × CBA/J)_{F₁} × C57BL/6J matings were transferred in 10μ l (10^5 IU) LIF to either day-4 pseudopregnant wild-type or pseudopregnant LIF-deficient homozygotes. After transfer the recipient was analysed for the presence of implantation sites on day 7.

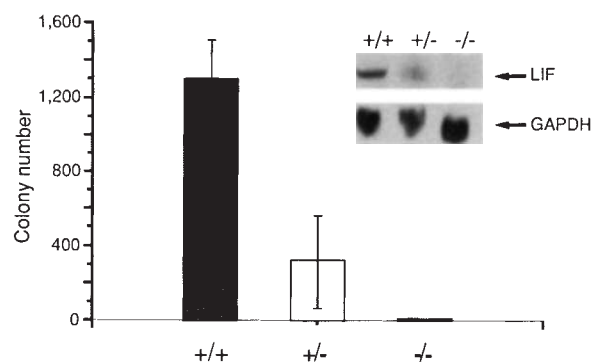


FIG. 3 The ability of PMEFs from wild-type, heterozygous and homozygous 13-day embryos to support the growth of undifferentiated wild-type ES cells was assayed. PMEFs (10^6) from embryos of each genotype were treated with mitomycin C²⁹ and plated onto 60-mm dishes. Two days later, 3×10^3 W9.3 ES cells were seeded onto each dish. Six days later, the ES colonies were analysed and counted. Each experiment was done in triplicate. Total RNA was prepared from the three PMEF lines and 25 μ g from each was analysed by northern blotting³⁰. The characteristic murine 4.2-kb LIF band was detected in the wild-type PMEFs, whereas the heterozygotes had a similar-sized band of reduced intensity, and in the homozygotes the band was absent. The integrity and amount of RNA loaded was determined using a glyceraldehyde phosphate dehydrogenase (GAPDH) probe.

complex^{18,19}. Thus, an absolute dependence on LIF for some physiological function might be circumvented if oncostatin M/CNTF or some other factor can act as a substitute. Although the existence of murine oncostatin M has yet to be demonstrated, our results show that there is an absolute requirement for LIF for implantation and also that ES cells are maintained in culture by PMEFs.

Cytokines have key roles in fetal-maternal relationships during pregnancy²⁰. For example, colony stimulating factor-1 (CSF-1) and its receptors are expressed at the fetal-maternal interface^{21,22}, and mutations in the maternal CSF-1 gene can compromise normal embryonic development, as can alterations

in the levels of other cytokines^{23,24}. Therefore cytokines, discovered largely because of their effects on the haemopoietic and immune systems, may be important in the regulation of reproduction and may be effectors²⁰ of endocrinological changes during mammalian pregnancy. LIF is another example of a cytokine with both embryonic and haemopoietic functions.

Here we have shown that the uterine expression of LIF in mice is essential for implantation, and it will be of interest to find out whether other mammals share this requirement. If so, LIF will not only provide a means to control pregnancy, but it may also help to improve the establishment of successful pregnancies from embryos produced by *in vitro* fertilization. □

Received 27 April; accepted 17 July 1992.

- Weitlauf, H. M. in *The Physiology of Reproduction* (eds Knobil, E. et al.) 231-261 (Raven, New York, 1988).
- Hodgen, G. D. & Itskovitz, J. in *The Physiology of Reproduction* (eds Knobil, E. et al.) 1995-2021 (Raven, New York, 1988).
- Bhatt, H., Brunet, L. J. & Stewart, C. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11408-11412 (1991).
- Gearing, D. P. et al. *EMBO J.* **6**, 3995-4001 (1987).
- Thompson, S. et al. *Cell* **56**, 313-321 (1989).
- Zijlstra, M. et al. *Nature* **344**, 742-746 (1990).
- McMahon, A. P. & Bradley, A. *Cell* **62**, 1073-1085 (1990).
- Capecci, M. R. *Science* **244**, 1288-1292 (1989).
- Williams, R. L. et al. *Nature* **336**, 684-686 (1988).
- Smith, A. G. et al. *Nature* **336**, 688-690 (1988).
- Matsui, V. et al. *Nature* **353**, 750-752 (1991).
- Murray, R., Lee, F. & Chi, C. P. *Molec. cell. Biol.* **10**, 4953-4956 (1990).
- Conquet, F. & Brulet, F. *Molec. cell. Biol.* **10**, 3801-3805 (1990).
- Hilton, D. J., Wearing, P. M., Nicola, N. A. & Metcalf, D. J. *Cell Physiol.* **148**, 430-439 (1991).
- Hilton, D. J. *Trends Biochem.* **17**, 72-76 (1992).
- Bazan, J. F. *Neuron* **7**, 197-208 (1991).
- Rose, T. M. & Bruce, A. G. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8641-8645 (1991).

- Gearing, D. P. & Bruce, A. G. *New Biol.* **4**, 61-64 (1992).
- Ip, N. Y. et al. *Cell* **69**, 1-11 (1992).
- Pamper, S., Arcedi, R. J. & Pollard, J. W. *Bioessays* **13**, 535-540 (1991).
- Pollard, J. W. et al. *Nature* **390**, 484-486 (1987).
- Regenstreif, L. J. & Rossant, J. *Devl Biol.* **133**, 284-294 (1989).
- Pollard, J. W., Hunt, J. S., Jedrejczak, W. W. & Stanley, E. R. *Devl Biol.* **148**, 273-283 (1991).
- Tartakovsky, B. & Ben-Yair, E. *Devl Biol.* **146**, 345-352 (1991).
- Mansour, S. L., Thomas, K. R. & Capecci, M. R. *Nature* **336**, 348-352 (1988).
- Stahl, J. et al. *J. biol. Chem.* **265**, 8833-8841 (1990).
- Stewart, C. L., Schuetz, S., Vanek, M. & Wagner, E. F. *EMBO J.* **6**, 383-388 (1987).
- Koentgen, F. thesis, Univ. Freiburg, Germany (1992).
- Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 113-151 (IRL, Oxford, 1987).
- Robertson, E. J. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 71-112 (IRL, Oxford, 1987).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).

ACKNOWLEDGEMENTS. We thank A. Carey and A. McMahon for critical reading and comments on the manuscript and S. Perry for efficient typing of the manuscript.

Infection breaks T-cell tolerance

Martin Röcken*, Joseph F. Urban†
& Ethan M. Shevach*

* Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA
† Department of Agriculture, Agricultural Research Service, Livestock and Poultry Science Institute, Helminthic Diseases Laboratory, Beltsville, Maryland 20705, USA

CLONAL deletion or clonal anergy establish tolerance in T cells that bear potentially autoreactive antigen receptors¹⁻¹⁵. Here we report that concomitant infection with the nematode *Nippostrongylus brasiliensis* breaks an established T-cell tolerance induced by injection of mice with *Staphylococcus enterotoxin B* (SEB)^{16,17}. CD4⁺ T cells from SEB-tolerant mice did not produce either interleukin-2 or interleukin-4 when challenged *in vitro* with SEB. *N. brasiliensis* infection of SEB-primed animals resulted in a normal expansion of SEB-tolerant CD4⁺Vβ8⁺ T cells *in vivo* as well as an equivalent increase of SEB-reactive, interleukin-4-producing CD4⁺Vβ8⁺ T cells both in SEB-tolerant and in normal animals. Thus, infection with *N. brasiliensis* circumvented the tolerance established with SEB. Activation of anergic, potentially autoreactive CD4⁺ T cells by infectious agents seems to be a major pathway for the initiation of autoimmune diseases¹⁵. Our results suggest that infectious agents may break tolerance in potentially autoreactive CD4⁺ T cells by activation of alternative reaction pathways.

Exposure of newborn mice to superantigens, such as MIs^a or SEB induces intrathymic clonal deletion of distinct T-cell families bearing common T-cell receptor (TCR) variable β-chain (Vβ) domains^{1,2,18}. In contrast, injection of adult animals with either cells expressing MIs^a or SEB induces only partial clonal deletion and is primarily characterized by a state of unresponsiveness or anergy that closely reflects the state of self-tolerance to MIs observed in radiation chimaeras or transgenic mice^{7-13,16,17,19-21}. This anergic state is manifested both by

decreased interleukin-2 (IL-2) production and a failure to proliferate after rechallenge *in vitro* with the tolerizing superantigen. Although the induction of anergy by exposure of mice to superantigens is now an established model for tolerance induction in mature CD4⁺ T cells and reflects many aspects of tolerance to self antigens^{7-10,13,16,17,19-21}, it is not known whether anergy induced by SEB completely silences SEB-reactive CD4⁺ T cells (bearing the TCR Vβ 8.1, 2, 3 chains¹⁸) or whether reactivity toward the tolerizing antigen can be restored *in vivo*. Because infectious agents have been postulated to trigger autoimmune diseases by activation of anergic, potentially autoreactive CD4⁺ T cells¹⁵, we have analysed whether infection of mice with the nematode *N. brasiliensis*, which is a potent inducer of proliferation and lymphokine production by CD4⁺ T cells *in vivo*²², can reverse the tolerant state induced by injection of animals with SEB.

Mice were first rendered anergic by a single (or three consecutive) intravenous injections of SEB¹⁷ (100 μg per injection) and 5 days later (or 2 days after multiple injections) infected with *N. brasiliensis*. As polyclonal expansion of all lymphoid cells is a hallmark of *N. brasiliensis* infection²², the animals were killed on day 7 or 8 of infection (10-13 days after the injection of SEB), and the expansion of tolerized CD4⁺Vβ8⁺ cells was compared to the expansion of total mesenteric lymph node cells (LNC). Infection with *N. brasiliensis* resulted in a two- to fourfold expansion of total LNC irrespective of whether the animal has been previously treated with SEB (Table 1a, b). As described previously, the percentage of CD4⁺Vβ8⁺, but not of CD4⁺Vβ6⁺, cells was decreased in LNC of SEB-treated mice (Table 1a). But after infection with *N. brasiliensis* the degree of expansion of CD4⁺Vβ8⁺, CD4⁺Vβ6⁺ cells and total LNC was similar in SEB-treated mice and equivalent to the expansion of CD4⁺Vβ8⁺ or CD4⁺Vβ6⁺ subpopulations from animals infected with *N. brasiliensis* that were not treated with SEB (Table 1a, b). In contrast, rechallenge with SEB did not induce expansion of CD4⁺Vβ8⁺ T cells in SEB-treated mice (Fig. 1).

Although these data indicate that SEB-tolerant CD4⁺Vβ8⁺ T cells can proliferate *in vivo* in response to infection with

TABLE 1 Anergic CD4⁺Vβ8⁺ T cells proliferate normally during infection with *N. brasiliensis*(a) Absolute and relative numbers of mesenteric lymph node cells, CD4⁺Vβ6⁺ and CD4⁺Vβ8⁺ subpopulation

Treatment protocol			Total LNC (×10 ⁶)	CD4 ⁺ Vβ8 ⁺ (% of LNC)	Vβ8 ⁺ /CD4 ⁺	CD4 ⁺ Vβ6 ⁺ (% of LNC)	Vβ6 ⁺ /CD4 ⁺	Total CD4 ⁺ Vβ8 ⁺ (×10 ⁶ per animal)	Total CD4 ⁺ Vβ6 ⁺ (×10 ⁶ per animal)
Experiment 1	1 SEB	None	25	4.5	9.8	6.6	16.8	1.1	1.6
	2 SEB	<i>N. brasiliensis</i>	95	4.2	8.0	6.3	16.0	4.0	6.0
	3 None	<i>N. brasiliensis</i>	100	9.8	19.3	5.2	11.7	9.8	5.2
	4 None	None	34	11.0	18.6	6.4	13.7	3.7	2.2
Experiment 2	1 SEB	None	23	4.2	10.9	5.4	13.5	1.0	1.3
	2 SEB	<i>N. brasiliensis</i>	67	3.7	9.8	4.8	13.2	2.5	3.2
	3 None	<i>N. brasiliensis</i>	70	8.4	21.5	4.1	10.3	5.8	2.9
	4 None	None	23	10.4	19.3	4.9	9.4	2.4	1.1
Experiment 3	1 SEB	None	22	5.3	12.3	5.5	13.0	1.2	1.2
	2 SEB	<i>N. brasiliensis</i>	59	5.6	13.0	5.2	14.6	3.3	3.1
	3 None	<i>N. brasiliensis</i>	70	7.8	20.2	4.2	10.9	5.5	2.9
	4 None	None	30	8.8	20.1	5.1	11.1	2.6	1.5

(b) Relative increase of mesenteric lymph node cells, CD4Vβ6⁺ and CD4Vβ8⁺ subpopulations

		Mice tolerized with SEB		Control mice		
Relative increase after infection	LNC	CD4Vβ6 ⁺	CD4Vβ8 ⁺	LNC	CD4Vβ6 ⁺	CD4Vβ8 ⁺
Experiment 1	3.8×	3.8×	<u>3.6×</u>	2.9×	2.4×	2.6×
Experiment 2	2.9×	2.5×	<u>2.5×</u>	3.0×	2.6×	2.4×
Experiment 3	2.7×	2.6×	<u>2.8×</u>	2.3×	1.9×	2.1×

Mesenteric LNC were isolated from animals treated as in Fig. 2. Animals in experiment 3 received 3 injections of SEB and were killed on day 7 after infection. Viable cell counts were determined and single-cell suspensions (1×10^6 per tube) were labelled for two colour flow cytometry as described²⁶ with optimal dilutions of phycoerythrin-conjugated anti-CD4 (Becton Dickinson) and fluorescein isothiocyanate-conjugated anti-Vβ8.1, 2 (MR5-2 mouse IgG2a), anti-Vβ6 (RR4-7; rat IgG2b), or, as a control, anti-IgG1 antibody (10.9; mouse IgG2a), all from Pharmingen. Dead cells were gated by forward scatter and propidium iodide and 6,000 events were analysed on a Becton-Dickinson FACScan.

TABLE 2 Frequency of SEB-reactive IL-2 or IL-4 producing CD4⁺Vβ8⁺ T lymphocytes(a) Frequency of IL-2-producing CD4⁺Vβ8⁺ lymphocytes

Treatment protocol		Experiment 1	Experiment 2	Experiment 3
1 SEB	None	1.1×10^{-3}	2.2×10^{-3}	0.4×10^{-3}
2 SEB	<i>N. brasiliensis</i>	0.6×10^{-3}	0.4×10^{-3}	1.0×10^{-3}
3 None	<i>N. brasiliensis</i>	7.5×10^{-3}	5.8×10^{-3}	8.8×10^{-3}
4 None	None	7.5×10^{-3}	33.7×10^{-3}	9.8×10^{-3}

(b) Frequency of IL-4-producing CD4⁺Vβ8⁺ lymphocytes

1 SEB	None	0.7×10^{-4}	1.5×10^{-4}	0.4×10^{-4}
2 SEB	<i>N. brasiliensis</i>	13.9×10^{-4}	27.1×10^{-4}	26.0×10^{-4}
3 None	<i>N. brasiliensis</i>	19.4×10^{-4}	75.0×10^{-4}	43.9×10^{-4}
4 None	None	0.7×10^{-4}	0.7×10^{-4}	0.3×10^{-4}

Mesenteric LNC were isolated from animals treated as for Fig. 2. Animals in experiment 3 received 3 injections of SEB. Cells (30×10^6) were labelled for two-colour cytometry as described²⁶ with predetermined dilutions of anti-CD8-PE (Pharmingen) and anti-Vβ8 antibody F23.1-FITC²⁷. CD8-F23.1⁺ cells were positively sorted either with a FACStar PLUS (Becton Dickinson) or an Epics V (Coulter Electronics) and were $\geq 90\%$ CD8-F23.1⁺ at reanalysis. To determine the precursor frequency of IL-2-producing cells 100, 25, 6.6 and 1.6 cells were activated for 2 days in 30 μ l culture medium with 500 T-depleted syngeneic spleen cells (3,000 rad) and $10 \mu\text{g ml}^{-1}$ SEB. To determine the precursor frequency of IL-4-producing cells, 1,000, 250, 66 and 16 cells were activated for 2 days in 30 μ l culture medium with 500 T-depleted syngeneic spleen cells (3,000 rad), 4 units ml^{-1} IL-2 and $10 \mu\text{g ml}^{-1}$ SEB. After 2 days plates containing >100 T cells per well were irradiated (1,000 rad) and 1,000 IL-2-responsive CTLL cells or 1,000 IL-4 responsive CT4S cells²⁸ were added as indicator cells. After 1 more day IL-2 was added to CTLL, IL-4 to CT4S and 6 h later the cultures were labelled overnight with 1 μCi (³H)-thymidine (Du Pont de Nemours). The cultures were collected²⁶ and the frequency of IL-2- or IL-4-producing cells was calculated using the weighted mean method. The cut-off for 'positive' cultures was arbitrarily set at c.p.m.-counts that were a least $>\text{mean} + 3 \text{ s.e.m.}$ of cultures containing T-depleted syngeneic spleen cells + SEB ($\pm$ IL-2) and indicator cells; control cultures containing anti-IL-4 antibody confirmed that CTLL determined the frequency for IL-2-producing cells. The frequency for IL-2 or IL-4 producers were determined in independent experiments.

N. brasiliensis, they do not address the question of whether infection reversed the state of unresponsiveness to the tolerizing antigen. We therefore stimulated CD4⁺ T cells with SEB *in vitro* and quantitated both IL-2 (Fig. 2a, b) and interleukin-4 (IL-4) (Fig. 2c, d) production as well as the frequency of IL-2- and IL-4-producing CD4⁺ T cells (Table 2). Animals that received SEB alone (protocol 1) produced neither IL-2 nor IL-4 when challenged *in vitro* with SEB (Fig. 2a-d). Infection with *N. brasiliensis* did not reverse the 10- to 100-fold reduction in IL-2 production in response to SEB or to anti-Vβ8 monoclonal antibody (Fig. 2a, b, protocol 2) or the ~ 10 -fold reduction in frequency of SEB-reactive IL-2-producing cells (Table 2a). The downregulation of the IL-2 production was tolerogen specific, because CD4⁺ T cells from tolerized animals responded normally to *Staphylococcus enterotoxin A* (SEA; Fig. 2b).

After infection with *N. brasiliensis*, CD4⁺ T cells of SEB-treated animals could produce large amounts of IL-4 in response to the tolerizing antigen (up to 20,000 U IL-4 per 10^6 CD4⁺Vβ8⁺ cells, Fig. 2c, d, protocol 2), whereas uninfected mice released no detectable IL-4 when challenged with SEB *in vitro* (≤ 30 U IL-4 per 10^6 CD4⁺Vβ8⁺ cells; Fig. 2c, d, protocols 1 and 4). Because *in vitro* production of IL-4 is dependent on IL-2 (ref. 23), the addition of exogenous IL-2 was required for optimal IL-4 production by CD4⁺Vβ8⁺ cells from tolerant, but not from normal infected animals (Fig. 2c, protocols 2 and 3); however, addition of exogenous IL-2 could not induce IL-4 production by CD4⁺ T cells from uninfected animals (Fig. 2c, protocols 1 and 4). CD4⁺Vβ8⁺ cells from SEB-treated, infected animals also showed a normal (≥ 10 -fold) increase in IL-4 production when stimulated with anti-Vβ8 antibody (Fig. 2c). Similarly, after infection CD4⁺ T cells from SEB-treated and non-treated animals released equivalent amounts of IL-4 when stimulated with SEA (Fig. 2c, d). These results with bulk populations were supported by limiting dilution experiments that demonstrated a 20- to 50-fold increase of IL-4-producing, SEB-reactive CD4⁺Vβ8⁺ cells in populations derived from SEB-treated animals after infection (Table 2).

CD4⁺ T cells from mice treated with one single or three

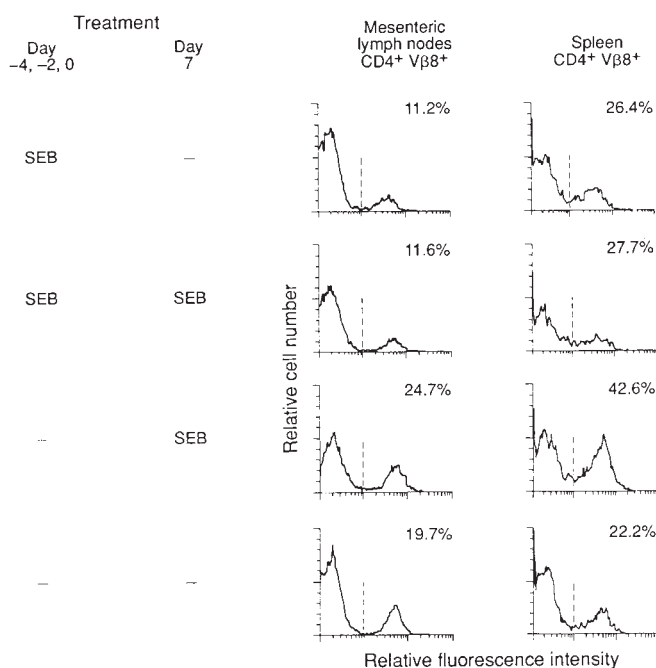


FIG. 1 Challenge with SEB does not induce expansion of CD4⁺Vβ8⁺ T cells in SEB-treated animals. The histograms show the percentage of CD4⁺Vβ8⁺ cells among CD4⁺ cells.

METHODS. Animals were injected with 100 µg SEB at the days indicated. On day 10, single-cell suspensions of mesenteric LNC or spleen cells (1×10^6 per tube) were labelled for two-colour flow cytometry as described²⁶ with optimal dilutions of phycoerythrin-conjugated (PE) anti-CD4 (Becton Dickinson) and fluorescein (FITC)-conjugated anti-Vβ8.1, 2, or, as a control, anti-IgG1 antibody. Dead cells were gated by forward scatter and propidium iodide and 10,000 events were analysed on a Becton-Dickinson FACScan. Each figure shown represents results of one of three separate animals with a similar cell distribution.

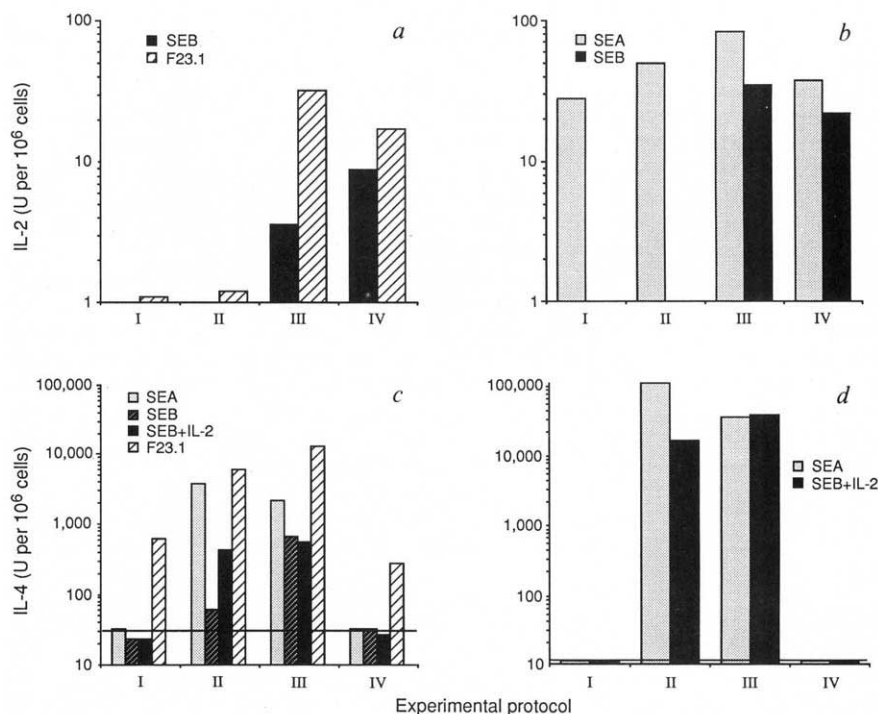
consecutive injections of SEB failed to produce IL-2 when challenged *in vitro* with either SEB or anti-Vβ8, even after infection (Fig. 2a, b and Table 2). Furthermore, even multiple injections of SEB alone did not induce a detectable population of SEB-reactive IL-4-producing CD4⁺ T cells (Fig. 2c, d and Table 2). Taken together, these results render it highly unlikely that the 20- to 50-fold increase in SEB-specific IL-4 producers

resulted from the expansion of a minor, nondetectable subset of SEB-reactive CD4⁺Vβ8⁺ cells.

These results offer new insights to our understanding of the anergic state of CD4⁺ T cells following injection of superantigens. Although SEB-tolerant CD4⁺ T cells demonstrate a profound downregulation of their ability to produce IL-2 *in vitro*, their capacity to expand and to produce IL-4 *in vivo* following

FIG. 2 *N. brasiliensis* infection fails to restore IL-2 production by SEB-tolerant CD4⁺ T cells to stimulation by SEB or anti-Vβ8 (a, b), but induces SEB-reactive IL-4-producing CD4⁺ T cells in animals tolerized with SEB (c, d). Each experiment included animal groups treated with one of the four protocols: 1, SEB-tolerized; 2, SEB-tolerized, *N. brasiliensis* infected; 3, non-tolerized, infected; 4, no treatment. Similar results were derived from six independent experiments.

METHODS. Female BALB/c mice (6–12 weeks of age) were injected with SEB (100 µg, Sigma) into the tail vein once (a, c) or 3 times (days -4, -2, 0; b, d). Five days later (or 2 days after multiple injections), the mice were inoculated with 700 third stage larvae as previously described²². After an additional 7–8 days, all animals were killed, single-cell suspensions prepared, and enriched populations of CD4⁺ T cells isolated with anti-CD8 labelled mouse T-cell columns (Biotex, Canada). CD4⁺ cells (0.1 – 0.5×10^6 per well) were cultured for 2 (a, c, d) or 3 (b) days in the presence of 3,000 rad irradiated T-depleted syngeneic spleen cells (0.2×10^6) in 250 µl medium (Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 5×10^{-5} M 2-mercaptoethanol, HEPES buffer, glutamine, penicillin and gentamycin) containing SEB (black columns; $10 \mu\text{g ml}^{-1}$), SEA (stippled columns; $25 \mu\text{g ml}^{-1}$; Toxin Technology) or with coated anti-Vβ8 (clear striped columns; antibody F23.1, $3 \mu\text{g ml}^{-1}$). For IL-4 production recombinant IL-2 (4 units ml^{-1} ; Cetus) was added together with SEB, except in c (dark striped columns). The maximal IL-2 release into the culture supernatants (by cultures containing 0.1 or 0.5×10^6 cells) was determined²⁶ by assay on IL-2-responsive CTLL cells (in the presence of $10 \mu\text{g ml}^{-1}$ anti-IL-4 antibody 11B11 (ref. 29). The IL-4 content of the culture supernatants was determined by assay



on only IL-4-responsive CT4S cells²⁸ as described^{23,26}. The data are expressed as Units (U) per 10^6 CD4⁺Vβ8⁺ T cells (for stimulations with SEB or antibody F23.1) or CD4⁺Vβ6⁺ T cells (for stimulations with SEA). The detection limit for IL-2 was <0.5 U per 10^6 CD4⁺Vβ8⁺ T cells (a, b), the detection limit for IL-4 is indicated by lines (c, d). Each panel was derived from an independent experiment.

infection with *N. brasiliensis* or injection of anti-IgD (M.R., F. D. Finkelman and E.M.S., unpublished data) seems to be preserved. Because IL-4-producing CD4⁺ T cells have been shown to be associated with various experimentally induced autoimmune diseases²⁴ and because many autoimmune diseases

(in man as well as in experimental animals) are associated with preceding infections^{15,25}, our results are consistent with a model for the development of autoimmune diseases in which infectious agents reverse the state of anergy of CD4⁺ T cells tolerant to self-antigens by activation of alternative reaction pathways. □

Received 1 June; accepted 14 July 1992.

- Kappler, J. W., Staerz, U., White, J. & Marrack, P. C. *Nature* **332**, 35–40 (1988).
- MacDonald, H. R. *et al. Nature* **332**, 40–45 (1988).
- Kisielow, P., Blüthmann, H., Staerz, U. D., Steinmetz, M. & von Boehmer, H. *Nature* **333**, 742–746 (1988).
- Smith, H., Chen, I.-M., Kubo, R. & Tung, K. S. K. *Science* **245**, 749–752 (1989).
- Ramsdell, F., Lantz, T. & Fowlkes, B. J. *Science* **246**, 1038–1041 (1989).
- Fry, A. M., Jones, L. A., Kruisbeek, A. M. & Matis, L. A. *Science* **246**, 1044–1046 (1989).
- Hodes, R. J., Sharrow, S. O. & Solomon, A. *Science* **246**, 1041–1044 (1989).
- Blackman, M. A. *et al. Nature* **345**, 540–542 (1990).
- Yui, K., Komori, S., Katsumata, M., Siegel, R. M. & Greene, M. I. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7135–7139 (1990).
- Speiser, D. E., Chvatchko, Y., Zinkernagel, R. M. & MacDonald, H. R. *J. exp. Med.* **172**, 1305–1314 (1990).
- Burkly, L. C., Lo, D., Kanagawa, O., Brinster, R. L. & Flavell, R. A. *Nature* **342**, 564–566 (1989).
- Murphy, K. M., Weaver, C. T., Elish, M., Allen, P. M. & Loh, D. Y. *Proc. natn. Acad. Sci. U.S.A.* **86**, 10034–10038 (1989).
- Schwartz, R. H. *Science* **248**, 1349–1356 (1990).
- Sprent, J., Gao, E.-K. & Webb, S. R. *Science* **248**, 1357–1363 (1990).
- Sinha, S. A., Lopez, M. T. & McDevitt, H. O. *Science* **248**, 1380–1388 (1990).

- Relahan, B. L., Jones, L. A., Kruisbeek, A. M., Fry, A. M. & Matis, L. A. *J. exp. Med.* **172**, 1091–1100 (1990).
- Kawabe, Y. & Ochi, A. *Nature* **349**, 245–248 (1991).
- White, J. *et al. Cell* **56**, 27–35 (1989).
- Qin, S., Cobbold, S., Benjamin, R. & Waldmann, H. *J. exp. Med.* **169**, 779–794 (1989).
- Rammensee, H.-G., Kroschewski, R. & Frangoulis, B. *Nature* **339**, 541–544 (1989).
- Webb, S., Morris, C. & Sprent, J. *Cell* **63**, 1249–1256 (1990).
- Finkelman, F. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 9675–9678 (1986).
- Ben-Sasson, S. Z., Le Gros, G., Conrad, D. H., Finkelman, F. D. & Paul, W. E. *J. Immun.* **145**, 1127–1136 (1990).
- Goldman, M., Druet, P. & Gleichmann, E. *Immun. Today* **12**, 223–227 (1991).
- Ohashi, P. S. *et al. Cell* **65**, 305–317 (1991).
- Röcken, M. *et al. J. Immun.* **148**, 47–54 (1992).
- Staerz, U. D., Rammensee, H.-G., Benedetto, J. D. & Bevan, M. J. *Immun.* **134**, 3994–4000 (1985).
- Hu-Li, J., Ohara, J., Watson, C., Tsang, W. & Paul, W. E. *J. Immun.* **142**, 800–807 (1989).
- Ohara, J. & Paul, W. E. *Nature* **315**, 333–336 (1985).

ACKNOWLEDGEMENTS. We thank W. E. Paul and R. H. Schwartz for reading the manuscript, D. W. Alling for help with statistics, B. Classen, F. D. Finkelman and W. E. Paul for antibodies and cell lines, S. Barbieri and D. Stephany for cell sorting and S. Starnes for assistance with the manuscript.

Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage

L. David Sibley* & John C. Boothroyd†

Department of Microbiology and Immunology,
Stanford University School of Medicine,
Stanford, California 94305–5402, USA

THE protozoan *Toxoplasma gondii* is a prevalent parasite in wild and domestic animals worldwide, being transmitted through the food chain by carnivorous feeding and scavenging¹. *Toxoplasma* normally divides asexually to yield a haploid form that can infect virtually any vertebrate but it also has a well defined sexual cycle that occurs exclusively in cats¹. *Toxoplasma* has become important as an often fatal opportunistic pathogen in patients with AIDS^{2,3}, although the 15–85% of adult human populations that are chronically infected with *T. gondii* are typically asymptomatic^{4–7}. Infections in immunocompromised hosts have variable outcomes. For example, only 30 to 50% of AIDS patients that are chronically infected with the parasite develop toxoplasmic encephalitis^{2,3} and only about half of acute maternal infections result in congenital disease of the newborn⁸. *T. gondii* strains differ in their virulence in animals¹, but the extent to which different strains are related has not been determined. Here we analyse 28 strains from a variety of hosts on five continents and find that the ten virulent strains have an essentially identical genotype, whereas the nonvirulent strains are moderately polymorphic. These data strongly suggest that virulent strains of *T. gondii* originated from a single lineage which has remained genetically homogeneous despite being globally widespread, and despite the ability of this organism to reproduce sexually.

To determine their genetic relatedness, we examined the genotype of *T. gondii* strains from a variety of hosts throughout the world (Table 1). These strains form two distinct groups based on their virulence in mice during acute infection: those that are highly virulent, and those that are nonvirulent and give rise to chronic infections (Table 1). Strains were analysed with a subset of gene probes randomly selected from a large set of

markers characterized for use in genetic and physical mapping of the *Toxoplasma* genome^{9,10}. The first probe, SAG1, is a single-copy gene encoding the major surface antigen, p30 (ref. 11); the second, probe 850, is also derived from a single-copy gene but its coding function is unknown^{9,10}; the third, probe BS, is derived from a repetitive element dispersed in the *Toxoplasma* genome and is also of unknown coding function¹⁰. Sufficient sequence is known for these three probes that the polymerase chain reaction (PCR; ref. 12) could be used to amplify small segments of genomic DNA for analysis of restriction fragment length polymorphisms (RFLPs), enabling isolates to be analysed with little or no growth outside the original host, thereby minimizing the chance of selecting for or against a given genotype.

A sample of each strain was subjected to lysis, PCR amplification and RFLP analysis. When strains were analysed for three polymorphic restriction sites in the SAG1 locus, all virulent strains gave the same RFLP pattern (allele) which was distinct from the allele seen in all nonvirulent strains (Fig. 1a, and Table 1). The second representative locus, hybridizing to probe 850, also had two alleles, with all virulent strains giving the same RFLP pattern (Fig. 1b, and Table 1) and most non-virulent strains showing a second, distinct allele. But in this case, several of the nonvirulent strains (including CEP and Tg51; Fig. 1b) showed the 'virulent' RFLP pattern at the 850 locus. The BS probe also gave essentially the same RFLP pattern for all virulent strains, whereas the nonvirulent strain patterns were all dissimilar to that from the virulent strains (Fig. 2a and Table 1; other data not shown).

We used seven other probes to investigate a representative subset of 11 strains by direct RFLP analysis of total genomic DNA, without PCR amplification. These probes were randomly selected from a large set of single-copy markers, the only criterion being that they should be genetically unlinked^{9,10}. They included two complementary DNAs of known coding function, three cDNAs of unknown coding function, and two randomly selected genomic fragments that were probably from non-coding regions. The results are summarized in Table 2 and confirm the conclusions based on the results with the three PCR probes: virulent strains have indistinguishable genotypes and are distinct from nonvirulent strains, which are themselves heterogeneous.

The multiple-copy BS probe was also used for RFLP analysis of total genomic DNA. The results (Fig. 2b) show that this probe can serve to 'fingerprint' strains, as virtually all lanes show a distinct pattern. The exceptions to this include the two Brazilian isolates OH3 and S11, which came from the same

* Present address: Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, Missouri 63110, USA.

† To whom correspondence should be addressed.

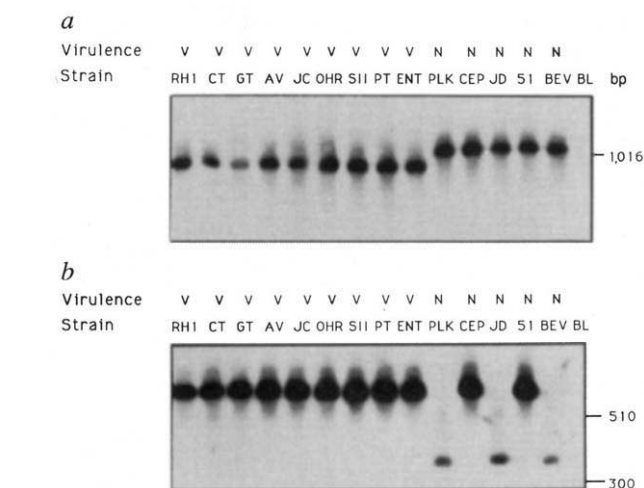
FIG. 1 *a*, Correlation of *Toxoplasma* virulence with genotypes as assessed by PCR/RFLP patterns. The SAG1 locus was amplified by PCR, digested with *DdeI* and analysed by Southern blotting using a fragment of the SAG1 gene as probe. All nine virulent (V) strains share the same allele characterized by a smaller restriction fragment than that seen in the five nonvirulent (N) strains. Similar results were obtained with this probe after using *HaeIII* and *Sau96I* for digestion (there was a perfect correlation between virulence and presence of a *Sau96I* site and absence of a *HaeIII* site; data not shown). The complete sequence of the SAG1 gene from three of these strains (RH1, CEP and PLK) is known²¹; the allele present in the two nonvirulent strains, CEP and PLK, differed by only a single nucleotide out of 1,142 (not affecting a restriction site). The allele from the virulent strain RH differed from CEP and PLK at 14 and 15 positions, respectively, including the three polymorphic restriction sites examined here. *b*, The 850 locus was analysed by PCR/RFLP as described for *a* but using *RsaI* digestion and the 850 probe. Whereas all nine virulent strains show the same genotype, only three of the five nonvirulent strains have a different allele, with the remaining two apparently having the 'virulent' type of allele. Similar results were obtained after *Sau96I* or *DdeI* digestion (the PCR product from all virulent and the same two nonvirulent strains shows a diagnostic 220-base-pair (bp) fragment with *Sau96I* and a 300-bp fragment with *DdeI*; the remaining three nonvirulent strains give 200-bp and 340-bp fragments, respectively, with these two enzymes). Strains are listed in Table 1; BL refers to the PCR blank, or negative control.

METHODS. *Toxoplasma* cells, grown in human fibroblasts, purified as described^{9,10}, were resuspended in phosphate-buffered saline and heated to 90 °C for 5 min before storage at -20 °C as a cell lysate. A sample of this lysate representing ~10⁴ cells was amplified by PCR as described²⁴ with the following adaptations. For SAG1, the two primers were 5'-CAACGGTAATCACTCAGCG-3' and 5'-CAATGTGCACCTGTAGGAAGC-3' and each of the 40 reaction cycles consisted of 95 °C for 1 min, 60 °C for 1 min

and 72 °C for 1 min. For 850, oligonucleotides were 5'-AAGACCTGGT-AACAGTCC-3' and 5'-TCAAGGCTTGGATGTTTCG-3' and each of the 40 cycles consisted of 95 °C for 1 min, 50 °C for 2 mins and 72 °C for 2 mins. PCR products were purified by phenol/chloroform extraction, digested with restriction endonuclease, analysed by agarose gel electrophoresis and blotting to Nytran membranes (Schleicher and Schuell)^{9,10}. α -³²P-radiolabelled probes consisted of the cloned gene fragments derived from RH strain material: probe SAG1 was the full-length gene (the AVA-1 fragment of ref. 11); probe 850 was an 850-bp *EcoRI* fragment encompassing the PCR region. Sizes in base pairs are shown for standards run in parallel.

geographic region and may be epidemiologically related (P. D. Glasner, personal communication). The unusual degree of polymorphism detected in genomic blots with probe BS contrasts with the results with single-copy probes (those used here and other unpublished results) and is consistent with the dispersed, repetitive nature of this probe. Thus, BS provides an important control showing that the similarity of genotypes among virulent strains is not caused by laboratory contamination.

From the data in Tables 1 and 2, we draw the following conclusions. First, for at least two loci, only two alleles are detected in the total gene pool of *T. gondii* worldwide. Second, there is no apparent correlation between genotype and host or geographic location from which the isolate was obtained. Third, with one exception (strain JC and the c19 probe), all virulent



strains show the same RFLP pattern when low-copy-number genes are analysed, indicating that they are genetically homogeneous. The polymorphism seen with the multiple-copy BS probe in genomic Southern blots does not conflict with our conclusion of homogeneity from results with the single-copy probes, as even otherwise identical genotypes would show variation in hyper-polymorphic sequences such as dispersed repetitive elements.

The human isolates examined here by RFLP analysis (both virulent and nonvirulent in mice) are not of a single genotype, a result that is consistent with an isoenzyme study of a collection of primarily French, human isolates which comprised five distinct zymodemes^{13,14}. The lack of similarity among human *T. gondii* isolates is not typical of other pathogens responsible for

strains show the same RFLP pattern when low-copy-number genes are analysed, indicating that they are genetically homogeneous. The polymorphism seen with the multiple-copy BS probe in genomic Southern blots does not conflict with our conclusion of homogeneity from results with the single-copy probes, as even otherwise identical genotypes would show variation in hyper-polymorphic sequences such as dispersed repetitive elements.

The human isolates examined here by RFLP analysis (both virulent and nonvirulent in mice) are not of a single genotype, a result that is consistent with an isoenzyme study of a collection of primarily French, human isolates which comprised five distinct zymodemes^{13,14}. The lack of similarity among human *T. gondii* isolates is not typical of other pathogens responsible for

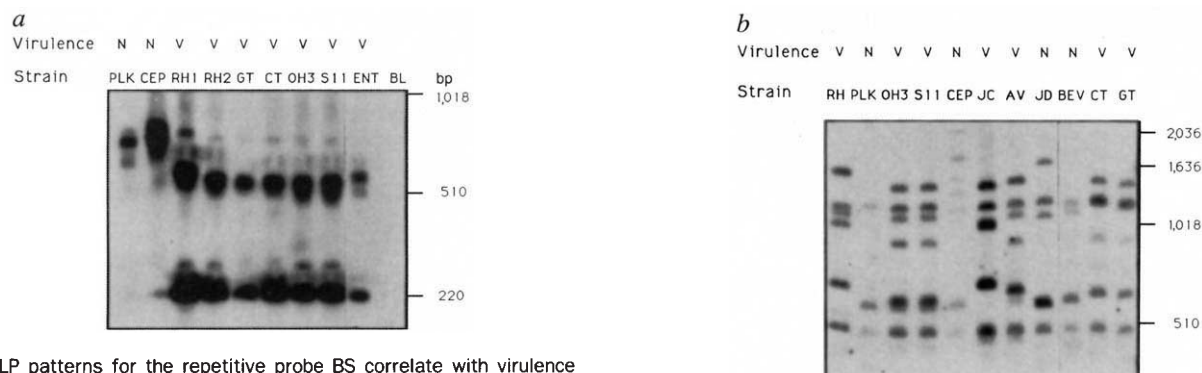


FIG. 2 RFLP patterns for the repetitive probe BS correlate with virulence when analysed by PCR/RFLP, but serve as a 'fingerprint' when analysed by genomic Southern blot. *a*, The BS repeats in genomic DNA from nine strains were amplified by PCR, digested with *MspI* and analysed by Southern blotting using a fragment of one BS repeat as probe. All seven virulent (V) strains show essentially the same pattern (designated allele 1 in Table 1). In contrast, two representative nonvirulent strains show distinct alleles (in this exposure, the signal from the two major bands for strain CEP have coalesced). *b*, Total genomic DNA from 11 strains (including the nine analysed in *a*) was digested with *MspI* and analysed by Southern blotting using the BS probe. Among the seven virulent strains there are four strains that fall into two pairs based on their almost identical patterns (OH3/S11 and CT/GT). The remainder of the virulent and nonvirulent strains give highly individual

patterns. Sizes in base pairs are shown for standards run in parallel.

METHODS. *a*, PCR/RFLP analysis was as described in Fig. 1, except that the primers used were 5'-CAGCCTGTCTATGGTTCC-3' and 5'-AGAAGAAA-CGGCAAGGGTCC-3', digestion was with *MspI*, and the probe was the BS probe cloned from RH1 strain. This probe was originally identified as a 700-bp repeat defined by *BamHI* digestion of a randomly isolated cosmid, c10. *b*, About 2.0 µg of *Toxoplasma* genomic DNA prepared as described^{9,10} from each of the 11 strains was digested with *MspI* and resolved on a 1.5% agarose gel, transferred to a Nytran membrane, and hybridized to the same probe as in *a*. The RH sample is from RH1.

TABLE 1 Correlation of virulence with genetic markers in *Toxoplasma* strains

Strain	Host	Source	Location	Virulence	SAG1	850	Genotypes at three loci		
							BS		
							<i>MspI</i>	<i>DdeI</i>	<i>HhaI</i>
CT1	Cow	JD	USA	+	1	1	1	1	1
GT1	Goat	JD	USA	+	1	1	1	1	1
S11	Pig	MM	Brazil	+	1	1	1	1	1
RH1*	Human	EP	USA	+	1	1	1	1	1
RH2	Human	JR	USA	+	1	1	1	1	1
OH3	Human	MM	Brazil	+	1	1	1	1	1
PT	Human	MD	France	+	1	1	1	1	1
ENT	Human	MD	France	+	1	1	1	1	1
AV	Human (AIDS)	JR/BD	USA	+	1	1	1	1	1
JC	Human (AIDS)	JR	USA	+	1	1	1	1	1
ME49	Sheep	JR	USA	-	2	2	2	2	2
PLK*	Sheep	LK	USA	-	2	2	2	2	2
CEP*	Cat	EP	USA	-	2	1	3	3	3
M7741	Sheep	EP	USA	-	2	1	4	3	3
C56	Chicken	JR	USA	-	2	1	4	3	3
Tg17	Pig	JF	Japan	-	2	2	2	2	2
S1	Sheep	MD	France	-	2	2	2	2	2
BEV	Rabbit	JF	UK	-	2	2	2	2	2
Tg51	Human	JF	France	-	2	1	3	3	3
Tg68	Human	JF	USA	-	2	2	3	3	3
Tg96	Human	JF	Australia	-	2	2	2	2	4
Tg132	Human	JF	Japan	-	2	2	5	2	5
WF	Human (AIDS)	JR	USA	-	2	2	6	3	2
MS	Human (AIDS)	JR	USA	-	2	2	6	3	2
AP	Human (AIDS)	JR	USA	-	2	2	7	4	3
WL	Human (AIDS)	JR	USA	-	2	2	8	5	2
HR	Human (AIDS)	JR	USA	-	2	1	6	3	3
JD	Human (AIDS)	JR	USA	-	2	2	2	2	2

Genotypes are reported based on the pattern of RFLPs detected from PCR-amplified material as described in figure legends. Numbers indicate like alleles. For the SAG1 locus^{11,21} genotype 1 signifies the allele having sites for *DdeI* and *Sau96I* at positions +751 and -100, respectively, and lacking a *HaeIII* site at position +133. Allele 2 of this gene lacks the *DdeI* and *Sau96I* sites but has the *HaeIII* site. For the 850 locus, allele designation is given in the legend to Fig. 1b. Strains: PT refers to strain 'P' isolated in Toulouse, France (Darde) which we have renamed to prevent confusion with other strains named P. Likewise, the original P strain of Kaspar²² has been renamed PLK, and the original C strain of Pfefferkorn²⁰ has been renamed CEP. RH1 and RH2 are clonal and nonclonal lines, respectively, of the original RH isolate²³ which have been passaged in different laboratories for over 20 years. Virulence definition was based on the 100% lethal dose (LD₁₀₀) of the strain in outbred Swiss mice, examined as soon as possible after original isolation (using organisms that had undergone fewest possible passages). 'Virulent' (+) relates to organisms with LD₁₀₀ < 10 and 'non-virulent' (-) to organisms with LD₁₀₀ > 10³. This distinction is based on published^{1,13,14,20,22,23} and unpublished information (from collaborators who provided strains and our own data). In some cases, the virulence of *Toxoplasma* strains increased somewhat after repeated passaging in the laboratory, but we have never observed a transformation from the nonvirulent to the virulent genotype. Virulence was not correlated with the frequency of laboratory passage in the majority of strains. *Toxoplasma* strains were obtained from the following labs: BD, B. Danneman, Rocky Mountain Laboratories, Hamilton, MT, USA; MD, M. L. Darde, Hôpital Universitaire Dupuytren, Limoges, France; JD, J. P. Dubey, Animal Parasitology Institute, USDA, Beltsville, MD, USA; JF, J. Frenkel, University of Kansas Medical Center, Kansas City, KA, USA; LK, L. Kasper, Dartmouth Medical School, Hanover, NH, USA; MM, M. C. Martins, Escola Paulista de Medicina, São Paulo, Brazil; EP, E. Pfefferkorn, Dartmouth Medical School, Hanover, NH, USA; JR, J. Remington, Palo Alto Medical Research Foundation, and Stanford University School of Medicine, Stanford, CA, USA.

* Isolates that were experimentally cloned by limiting dilution.

TABLE 2 Genotypes at respective single or low-copy number loci for virulent and non-virulent *Toxoplasma* strains

Strain	Virulence	SAG1	850	ROP1	950	Loci	SAG2	226	62B	c19
		(VII)	(V)	(?)	(VI)	L328	(VIII)	(IV)	(IX)	(X)
		DSH*	RSD*	<i>Sau3A</i>	<i>DdeI</i>	<i>RsaI</i>	<i>HhaI</i>	<i>HhaI</i>	<i>MspI</i>	<i>BamHI</i>
RH	+	1	1	1	1	1	1	1	1	1
OH3	+	1	1	1	1	1	1	1	1	1
S11	+	1	1	1	1	1	1	1	1	1
JC	+	1	1	1	1	1	1	1	1	3
AV	+	1	1	1	1	1	1	1	1	1
CT	+	1	1	1	1	1	1	1	1	1
GT	+	1	1	1	1	1	1	1	1	1
PLK	-	2	2	2	2	2	2	1	2	1
CEP	-	2	1	2	2	2	1	2	1	2
JD	-	2	2	2	2	2	2	1	2	1
BEV	-	2	2	2	2	2	2	1	2	2

Numbers signify like genotypes at respective single or low copy loci. Virulence as defined in Table 1. The chromosomes to which the gene probes map are shown in parentheses. Probes SAG1, SAG2 and ROP1 are cDNAs of known coding function; 850, 950, 226 and 65B are cDNAs of unknown coding function; and L328 and c19 are randomly selected genomic fragments^{9,10}.

* DSH indicates that the restriction endonucleases *DdeI*, *Sau96I* and *HaeIII* have been used for DNA digestion; RSD refers to restriction endonucleases *RsaI*, *Sau96I* and *DdeI*.

opportunistic infections¹⁵⁻¹⁷. This is probably because of epidemiological differences that limit the opportunities for human-to-human infection with *Toxoplasma*. In the absence of cannibalism, intraspecies transmission of *T. gondii* occurs only in cats¹ or, rarely, by transplacental infection⁸.

Given that the probes used here were randomly selected, it is unlikely that they are directly involved in producing the virulence phenotype. Hence, the high degree of genetic similarity among virulent strains of *T. gondii* suggests that these strains form a distinct subgroup. To test this possibility, we used a binomial calculation¹⁵ to determine the probability of observing a set of virulent strains with the same multilocus genotype using the null hypothesis of random mating. The calculation gave the probability of obtaining identical genotypes for 10 virulent strains of 28 total strains (Table 1) as $P = 8.8 \times 10^{-6}$. Using the smaller number of strains in Table 2, for which a larger number of loci have been analysed, $P = 1.2 \times 10^{-5}$ for the seven loci that are unlinked. These probability values are very low, indicating that the virulent strains examined here are not in random mating equilibrium with the population as a whole. Nonvirulent strains also show allele frequencies that differ from expected values, although this pattern is much less pronounced than for virulent strains. It is evident from Tables 1 and 2 that several possible recombinant genotypes are not observed (most notably there are no virulent strains with alleles predominant in nonvirulent strains). These findings are consistent with clonal expansion of virulent strains of *T. gondii* from a single lineage. The broad host range of *T. gondii* and the absence of geographical barriers may have contributed to the global distribution of the virulent genotype.

Although clonality has been observed among parasitic protozoa, many of these do not have a definitive sexual phase in their life cycle^{15,18}. In *T. gondii*, clonality occurs despite a well established sexual cycle¹ and evidence that it has contributed to the overall genetic diversity of several of these strains¹⁹. *T. gondii* is also highly unusual in that all virulent strains belong to a single clonal lineage. The mechanisms by which this clonality has arisen and is maintained cannot be determined from our study, but may involve exclusive self-mating and/or asexual reproduction. The ability to conduct classical genetic crosses in *Toxoplasma*^{9,10,20} could be exploited to distinguish between these alternatives and to identify the loci contributing to virulence in toxoplasmosis.

The fact that virulent strains of *T. gondii* comprise a single clonal lineage, regardless of their host or geographic origin, has clinical and biological implications. Using PCR it should be possible to examine fresh isolates from pregnant women and AIDS patients to establish whether there is any correlation between disease manifestations and infecting strains. Our results unambiguously extend the phenomenon of clonality of virulent strains to a ubiquitous protozoan species which has a sexual cycle. □

18. Tibayrenc, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5129-5133 (1991).
19. Boothroyd, J. C., LeBlanc, A. J. & Sibley, L. D. in *Toxoplasmosis: NATO ASI Series* (Springer, New York, in the press).
20. Pfefferkorn, E. R., Pfefferkorn, L. C. & Colby, E. D. *J. Parasit.* **63**, 158-159 (1977).
21. Buelow, R. & Boothroyd, J. C. *J. Immun.* **147**, 3496-3500 (1991).
22. Kasper, L. H. & Ware, P. L. *J. clin. Invest.* **75**, 1570-1577 (1985).
23. Sabin, A. B. *J. Am. med. Assoc.* **116**, 801-807 (1941).
24. Grover, C. M., Thulliez, P., Remington, J. S. & Boothroyd, J. C. *J. clin. Microbiol.* **28**, 2297-2301 (1990).

ACKNOWLEDGEMENTS. We thank our colleagues who contributed strains and unpublished information on their virulence, F. Araujo, J. Fishback and A. Sher; M.-L. Darde for communication of unpublished data; S. Beverley, F. Ayala and M. Tibayrenc for advice and help with statistical analysis; E. Boroff, A. Brookes, R. Buelow, J. Chen, A. LeBlanc and U. Play for assistance in probe characterization; and K. Kim and D. Soldati-Favre for critically reading the manuscript. This work was supported by the NIH, the John D. and Catherine T. MacArthur Foundation, Merck Sharp and Dohme and the Burroughs Wellcome Fund.

ERRATUM

A protein kinase homologue controls phosphorylation of ganciclovir in human cytomegalovirus-infected cells

V. Sullivan, C. L. Talarico, S. C. Stanat, M. Davis, D. M. Coen & K. K. Biron

Nature **358**, 162-164 (1992)

IN two figures in the paper in the 9 July issue, areas of shading failed to show up on the printed page. Revised versions below show the correct shading. The published figure legends are unchanged.

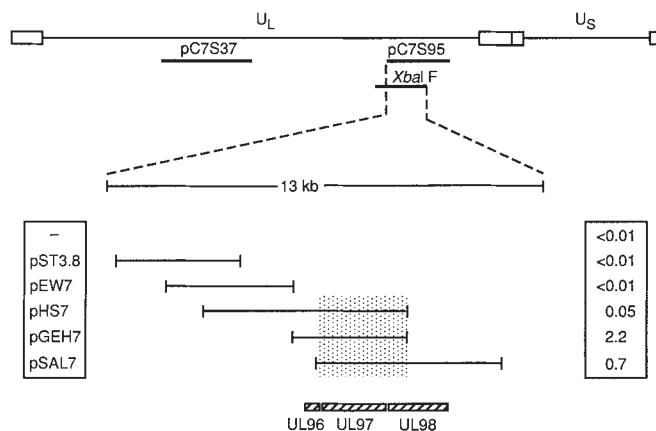


FIG. 1 Molecular mapping of ganciclovir-resistance markers of HCMV. The shaded box (lower centre) represents the 2.6-kb overlap (extending from *SalI* to *HindIII* sites) of plasmids pHS7, pGEH7 and pSAL7 which were able to transfer ganciclovir resistance.

	amino acid
HCMV UL97	622 - AspGluValArgMetGly- 9 - GlyAlaAlaCysArgAlaLeu
HHV-6 15R	429 - ArgGluAlaGlnLeuTyr- 9 - AspGluAlaCysArgLeuAsn
cAPK	220 - AspTrpTrpAlaLeuGly- 8 - GlyTyrProProPhePheAla

FIG. 3 A 4-amino-acid deletion in the HCMV UL97 gene product confers resistance to ganciclovir. The four amino acids deleted in ganciclovir-resistant mutant 759'D100 are shaded darkly and the five residues in subdomain IX of cAPK involved in substrate binding are lightly shaded.

Received 2 April; accepted 23 July 1992.

1. Dubey, J. P. in *Parasitic Protozoa* (ed. Kreier, J. P.) 101-237 (Academic, New York, 1977).
2. Luft, B. J. & Remington, J. S. *J. Inf. Dis.* **157**, 1-6 (1988).
3. Danneman, B. R. & Remington, J. S. *Hosp. Pract.* **24**, 139-154 (1989).
4. Hughes, H. P. A. *Parasit. Today* **1**, 41-44 (1985).
5. Desmonts, G. & Couvreur, J. *New Engl. J. Med.* **290**, 1110-1116 (1974).
6. Feldman, H. A. *Am. J. Epidemiol.* **81**, 385-391 (1965).
7. Apt, W. *Parasit. Today* **1**, 44-46 (1985).
8. Wilson, C. B. & Remington, J. S. *Am. J. Obstet. Gynec.* **138**, 357-363 (1980).
9. Sibley, L. D. & Boothroyd, J. C. *Molec. biochem. Parasitol.* **51**, 291-300 (1992).
10. Sibley, L. D., LeBlanc, A. J., Pfefferkorn, E. R. & Boothroyd, J. C. *Genetics* (in the press).
11. Burg, J. L., Perelman, D., Kasper, L. H., Ware, P. L. & Boothroyd, J. C. *J. Immun.* **141**, 3584-3591 (1988).
12. Saiki, R. K. et al. *Science* **230**, 1350-1354 (1985).
13. Darde, M. L., Bouteille, B. & Pestre-Alexandre, M. *Am. J. trop. Med. Hyg.* **39**, 551-558 (1988).
14. Darde, M. L., Bouteille, B. & Pestre-Alexandre, M. *J. Parasit.* (in the press).
15. Tibayrenc, M., Kjellberg, F. & Ayala, F. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2414-2418 (1990).
16. Selander, R. K. & Levin, B. R. *Science* **210**, 545-547 (1980).
17. Kiehn, T. E. et al. *J. clin. Microbiol.* **21**, 168-173 (1985).

Natural or artificially created mutant strains of mice are of great value in studying development and disease, but often such models reveal more differences than expected.

MOUSE models are in the news. The announcement¹ that mice homozygous for a defective version of the cystic fibrosis gene have been bred using homologous recombination is the latest example in which the manipulation of rodent genes has, or is expected to, cast light on development, gene regulation, immunology, memory or genetic disorders. Of course, more traditional transgenic animals and naturally occurring mutants have also proved to be of immense importance — in sex determination for example — but regardless of the origin of the mutation, quite often the model does not resemble the corresponding human condition as closely as might have been expected².

Mouse models for Lesch-Nyhan and Gaucher's disease have been created by gene targeting, but whereas the former seem to be asymptomatic, the latter — homozygous for a null allele — die within 24 hours of birth, making further constructs necessary before the more common presentations in humans can be mimicked. Mice containing disrupted alleles for p53 and the prion gene have also given unexpected phenotypes. A naturally occurring mutant is the *mdx* mouse, which contains a point mutation in the gene encoding dystrophin. Although a legitimate model for Duchenne muscular dystrophy (DMD), the *mdx* phenotype is far milder than DMD, with only the smooth muscle of the diaphragm showing signs of degeneration and fibrosis³. Furthermore, a cat model for muscular dystrophy⁴ shows muscle hypertrophy but not the myofibre leakage and progressive fibrosis seen in dystrophin-deficient humans or dogs.

In the September issue of *Nature Genetics*, another curious discrepancy is reported by Karen Steel and Richard Smith in their study⁵ of *Splootch* (*Sp/+*)

mice, which serve as a model for Waardenburg's syndrome type I (WSI). Waardenburg's syndrome type I is an autosomal dominant condition which accounts for some 2 per cent of cases of adult deafness. Other characteristics include facial irregularities and pigmentary abnormalities such as a white forelock and eyelashes. The severity of Waardenburg's syndrome is quite variable, and there are at least two distinct genetic loci. *Sp* is a pigmentary mutation in the *Pax-3* gene, a member of the paired-box class of homeobox genes. In the homozygous form, *Sp/Sp* mice suffer malformations of neural-crest-derived structures, including the inner ear and central nervous system. Three groups have documented mutations in individuals with WSI within the human homologue of *Sp*, *PAX3* (refs 6–8).

Given that hearing loss is a common feature of WS, it was expected that *Sp/+* mice would also be deaf, but surprisingly this is not the case⁵. Steel and Smith measured two important potentials, the endocochlear potential and the cochlear microphonics, generated by the organ of Corti and the outer hair cells, respectively. The recordings from *Sp/+* heterozygotes and wild types were indistinguishable. They also found that thresholds for the detection of low-sound stimuli (the compound action potential) in *Sp/+* heterozygotes are equal to (or slightly better than) in wild-type animals. Finally, the authors found no anatomical differences in the shape or size of the bony labyrinth or the middle ear⁵.

There are several possible reasons for the differences between humans and mice. First, the nature of the *Sp* mutant differs from the point mutations or small deletions found so far in human *PAX3*; *Sp*, by contrast, is a splice site mutation which removes the fourth exon of *Pax3*. Because the human mutations cause premature termination of the gene product, they may have more deleterious consequences than *Sp*. It is also possible that the penetrance of deafness in the mutant mice varies, as is evident in WS; indeed, genetic factors are known to influence the development of *Sp/Sp* fetuses. Similar analyses of other *Sp* alleles on a variety of genetic backgrounds might help to identify some of these modifying factors, which could aid the understanding of hearing loss in WS and neural crest defects in general.

The study points out that in some cases, however, mice and humans simply do not manifest the same phenotype despite very similar mutations^{2,5}. As has been seen for DMD and Lesch-Nyhan disease, mice often have milder phenotypes. Until recently, much the same was thought of another murine *Pax* gene mutation, *Small eye* (*Sey*), a semidominant, homozygous lethal mutation that affects the embryonic development of the nose and eyes and which serves as a model for human aniridia. In a paper documenting the first reported point mutations in *PAX6* in aniridia, Jordan *et al.*⁹ describe ocular defects in *Sey* heterozygotes more severe than in the past, underscoring the relationship between the two disorders.

Despite these phenotypic differences, mouse models continue to propel discoveries in human genetics. Two examples have emerged very recently. The human homologue of another mouse gene involved in pigmentation and development, the pink-eyed dilution (*p*) gene¹⁰, has been mapped to the region of chromosome 15q associated with Prader-Willi and Angelman syndromes. Two groups have also extended an earlier report by Tatsumi *et al.*¹¹ in detailing mutations in the *PIT-1* pituitary transcription factor which cause combined hormone deficiency and hypothyroidism^{12,13}. These studies were stimulated by the discovery of defects at the *Pit-1* locus in dwarf mice.

The observed differences between homologous defects in mice and humans may have a variety of genetic, anatomical and/or pathophysiological causes. Understanding these variations may prove to be as important as of the models themselves.

Kevin Davies

Kevin Davies is editor of Nature Genetics.

Also in this month's *Nature Genetics*: a specific DNA rearrangement on chromosome 4 associated with facioscapulohumeral muscular dystrophy; the creation of transgenic cystic fibrosis mice; mapping of genes contributing to autoimmune type I diabetes in the BB rat; linkage of a locus for a significant form of tuberous sclerosis near the polycystic kidney disease gene; and the physical relationship between a pair of imprinted genes in mice.

1. Snouwaert, J. N. *et al. Science* **257**, 1083–1088 (1992).
2. Collins, F. S. & Wilson, J. M. *Nature* **358**, 708–709 (1992).
3. Stedman, H. H. *et al. Nature* **352**, 536–539 (1991).
4. Gaschen, P. P. *et al. J. neuro. Sci.* **110**, 149–159 (1992).
5. Steel, K. P. & Smith, R. J. H. *Nature Genet.* **2**, 75–79 (1992).
6. Tassabehji, M. *et al. Nature* **355**, 635–636 (1992).
7. Baldwin, C. T. *et al. Nature* **355**, 637–638 (1992).
8. Morell, R. *et al. Hum. molec. Genet.* **1**, 243–247 (1992).
9. Jordan, T. *et al. Nature Genet.* **1**, 328–332 (1992).
10. Gardner, J. M. *et al. Science* **257**, 1121–1124 (1992).
11. Tatsumi, K. *et al. Nature Genet.* **1**, 56–58 (1992).
12. Radovick, S. *et al. Science* **257**, 1115–1118 (1992).
13. Präfle, R. W. *et al. Science* **257**, 1118–1121 (1992).

Reagents for the site-specific cleavage of megabase DNA

Peter B. Dervan

The physical mapping of chromosomes will be facilitated by methods of breaking large DNA into manageable fragments, or cutting uniquely at genetic markers of interest. Key issues in the design of sequence-specific DNA cleaving reagents are the specificity of binding, the number of different sequences that can be targeted and the cleavage yield.

OLIGONUCLEOTIDE-directed triple-helix formation is one of the most versatile methods for the sequence-specific recognition of double-helical DNA^{1,2}. The ability to target a broad range of DNA sequences, the high stabilities of the triple-helical complexes and the sensitivity to single base mismatches make this a powerful technique for binding single sites within large segments of double-helical DNA. This approach to DNA recognition has been used to mediate single site-specific cleavage of human chromosomal DNA³, as well as to inhibit transcription *in vitro*^{4,5}.

Recognition of double-helical DNA by triple-helix formation

At least two classes of DNA triple helices exist that differ in the sequence compositions of the third strand, relative orientations and positions of the backbones of the three strands, and base triplet interactions. Pyrimidine oligonucleotides bind purine tracts in the major groove of double-helical DNA parallel to the purine Watson-Crick strand. Sequence specificity is derived from thymine (T) recognition of adenine•thymine (AT) base pairs (T•AT base triplets) and N-3 protonated cytosine (C⁺) recognition of guanine•cytosine (GC) base pairs (C⁺•GC base triplets)⁶⁻⁸. An additional family of triple-helical structures consists of purine-rich oligonucleotides bound in the major groove to purine tracts of DNA antiparallel to the Watson-Crick purine strand^{4,9,10}. Sequence specificity is derived from guanine (G) recognition of GC base pairs (G•GC base triplets) and from A or T recognition of AT base pairs (A•AT and T•AT triplets)^{4,9,10}. The identification of other natural base triplets¹¹, alternate strand triple-helix formation^{12,13} and the design of non-natural base triplets¹⁴ have increased the general application of oligonucleotide-directed triple-helix formation for the recognition of DNA sequences containing all four base pairs.

The stability of triple helices is dependent on the length, sequence composition and modification of the oligonucleotide, as well as solution conditions, including pH and cation concentrations¹⁻¹⁵. A full understanding of the factors contributing to the stability and specificity of the binding of oligonucleotides to double-helical DNA will

require a complete characterization of the thermodynamics of the complex formation.

Oligonucleotides conjugated to EDTA•Fe can specifically bind double-helical DNA by triple-helix formation and produce double-strand cleavage at single sites in DNA¹. The free energy for oligodeoxynucleotide-directed triple-helix formation at a single site on a DNA plasmid fragment has been analysed using quantitative affinity cleavage titration¹⁵. Measurement of the amount of site-specific cleavage of a DNA restriction fragment (24 °C, 100 mM Na⁺, 1 mM spermine 4HCl, pH 7.0) over a concentration change of four orders of magnitude for oligodeoxynucleotide-EDTA•Fe (5'-T*TTTCTCTCTCTCT-3') yields an equilibrium binding constant, $K_T = 3.7 \times 10^6 \text{ M}^{-1}$ ($\Delta G_T = -9.0 \text{ kcal mol}^{-1}$) for a 15-base pair purine tract. Single internal base triplet mismatches result in a destabilization of the local triple-helical structure by 2.5–3.0 kcal mol⁻¹¹⁵.

Sequence-specific double-strand cleavage of DNA

The chemical yields for double-strand oxidative DNA cleavage by oligonucleotides equipped with EDTA•Fe are typically low (15–25%)¹. Such yields are adequate for addressing questions related to molecular recognition, including binding specificities, site size and orientation. If the full potential of oligonucleotide-directed recognition of double-helical DNA is to be realized for site-specific cleavage of DNA of high complexity, the design of oligonucleotides equipped with moieties capable of quantitative modification of DNA is an area for further development.

Enzymatic cleavage mediated by triple-helix formation. There appears to be two separate approaches that have in common coupling enzymatic cleavage with oligonucleotide-directed recognition of DNA. One is to attach DNA cleaving enzymes to oligonucleotides¹⁶. An oligonucleotide-staphylococcal nuclease adduct cleaves plasmid DNA in 75% yield¹⁶. Although the nuclease is, in a formal sense, nondiffusible, multiple cleavage sites at the binding site are observed, likely due to conformational flexibility of the attached

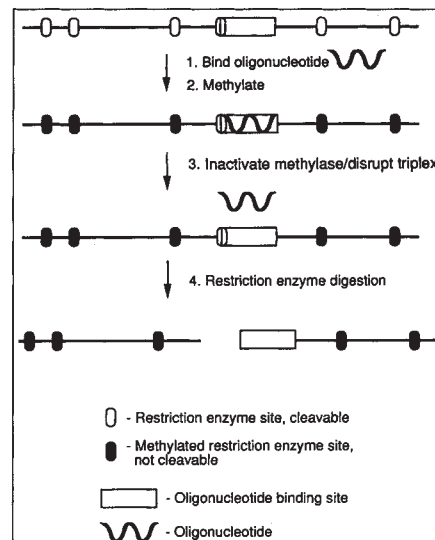


FIG. 1 General scheme for single-site enzymatic cleavage of genomic DNA by oligonucleotide-directed triple-helix formation. Chromosomal DNA is equilibrated with an oligonucleotide in a methylase-compatible buffer containing polycation. *EcoRI* methylase, which methylates the central adenines of the sequence 5'-GAATTC-3' and renders the sequence resistant to cleavage by *EcoRI* restriction endonuclease, is added and allowed to proceed to completion. The methylase is inactivated and the triple helix is disrupted at 55 °C in a high-pH buffer containing detergent. After washing extensively, the chromosomal DNA is re-equilibrated in restriction enzyme buffer and cut to completion with *EcoRI* restriction endonuclease. The cleavage products are separated by pulsed-field gel electrophoresis and efficiencies quantitated by Southern blotting (from ref. 18).

enzyme. The other approach is termed Achilles heel cleavage¹⁷ and involves transient site-specific protection from enzymatic methylation mediated by triple-helix formation, followed by triple-helix disruption and cleavage by a restriction enzyme¹⁸ (Fig. 1). Enzymatic cleavage mediated by this approach has been shown to cleave a single site in a yeast chromosome in 95% yield¹⁸ and human chromosome 4 (200 megabase pairs in size) in 80–90% yield³. Triple helix-mediated enzymatic cleavage affords high specificity that can, in principle, be customized to unique genetic markers without artificial insertion of a target sequence.

The use of degenerate oligonucleotides in this technique to screen rapidly genetic markers for overlapping triple-helix methylation–restriction sites could make it possible to cut chromosomal DNA uniquely and efficiently at endogenous sites with minimal sequence information³. It remains a challenge for chemists to match the enzymatic yields for DNA cleavage with rational mechanism-based chemical approaches for quantitative modification of DNA within a triple-helical complex.

Sequence-specific alkylation of double-helical DNA by triple-helix formation.

Attachment of the nondiffusible electrophile *N*-bromoacetyl to the 5' position of a thymine at the 5' end of a pyrimidine oligodeoxyribonucleotide affords sequence-specific alkylation of a guanine two base pairs to the 5' side of a local triple-helical complex in >96% yield¹⁹. *N*-bromoacetyl-oligodeoxyribonucleotides bind adjacent inverted purine tracts on double-helical DNA by triple-helix formation and alkylate single guanine positions on opposite strands at 37 °C (pH 7.4). After depurination, double-strand cleavage at a single site within plasmid DNA occurs in greater than 85% yield¹⁹ (Fig. 2). The resulting DNA fragments from site-specific alkylation and cleavage can be ligated with DNA fragments generated by restriction endonuclease digestion. This nonenzymatic approach, which couples sequence-specific recognition with sequence-dependent cleavage, affords double-strand, site-specific cleavage in megabase-size DNA. A yeast chromosome was cleaved at a single site in 85–90% yield¹⁹. This nonenzymatic approach affords double-strand cleavage yields comparable to those attained by triple helix-mediated enzymatic cleavage of DNA. Undoubtedly, the *N*-bromoacetyl-oligodeoxyribonucleotide-mediated cleavage is a first-generation approach and further refinements can be expected. Because oligonucleotide-directed recognition of double-helical DNA by triple-helix formation is one million times more specific than restriction enzymes, it is not unthinkable that quantitative reactions targeted to single atoms in the human genome (3 billion base pairs) should be possible by strictly chemical methods.

Current limitations

Oligonucleotide-directed targeting of double-helical DNA by triple-helix formation is still limited to mostly purine tracts^{1,2,9} or mixed sequences of the type (purine)_m(pyrimidine)_n and (pyrimidine)_m(purine)_n^{12,13}. Novel bases that bind TA and CG base pairs remain to be designed or discovered to complete the recognition code. For this reason, the recently reported RecA-mediated oligonucleotide Achilles heel procedure, which appears to allow all four base-pair targeting, shows significant promise for genome research²⁰.

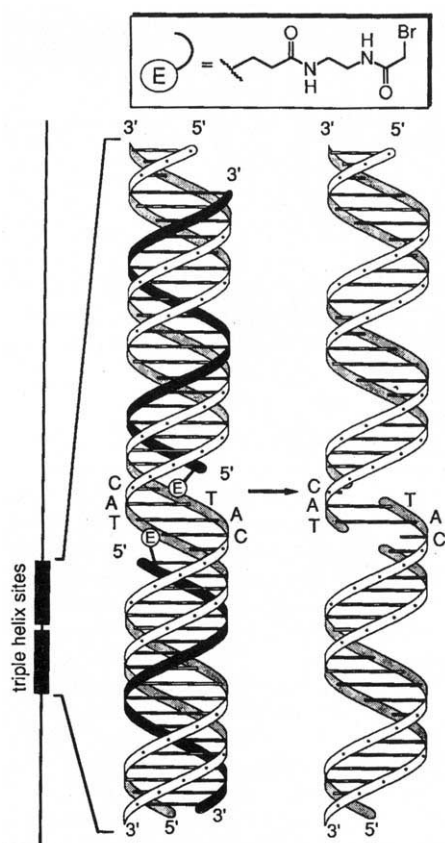


FIG. 2 Oligodeoxyribonucleotides equipped with an electrophile (E) at the 5' end bind to adjacent inverted binding sites on double-helical DNA by triple-helix formation and alkylate at single guanine positions on opposite strands¹⁹. After depurination, double-strand cleavage produces sequence-specific overhangs suitable for ligation.

With regard to single-site targeting of chromosomes *in vivo*, major technical hurdles still exist that will require both chemical and biological advances, such as (1) the design and efficient synthesis of modified oligonucleotide analogues that permeate cells and are resistant to nucleases and (2) a detailed understanding of both chromatin structure and transcriptional control *in vivo*. □

Peter B. Dervan is in the Division of Chemistry and Chemical Engineering at the California Institute of Technology, Pasadena, California 91125, USA. For more information, fill in reader service number 100.

1. Moser, H. E. & Dervan, P. B. *Science* **238**, 645–650 (1987).
2. LeDoan, T. et al. *Nucleic Acids Res.* **15**, 7749–7760 (1987).
3. Stobel, S. A., Doucette-Stamm, L. A., Riba, L., Housman, D. E. & Dervan, P. B. *Science* **254**, 1639–1642 (1991).
4. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, J. & Hogan, M. E. *Science* **241**, 456–459 (1989).
5. Maher, L. J., Dervan, P. B. & Wold, B. *Biochemistry* **31**, 70–81 (1992).
6. Feisenfeld, G., Davies, D. R. & Rich, A. *J. Am. chem. Soc.* **79**, 2023–2024 (1957).
7. Rajagopal, P. & Feigon, J. *Biochemistry* **28**, 7859–7870 (1989).

8. Radhakrishnan, I., Gao, X., de los Santos, C., Live, D. & Patel, D. J. *Biochemistry* **30**, 9022–9030 (1991).
9. Beal, P. A. & Dervan, P. B. *Science* **251**, 1360–1363 (1991).
10. Radhakrishnan, I., de los Santos, C. & Patel, D. J. *J. molec. Biol.* **221**, 1403–1418 (1991).
11. Griffin, L. C. & Dervan, P. B. *Science* **245**, 967–971 (1989).
12. Home, D. A. & Dervan, P. B. *J. Am. chem. Soc.* **112**, 2435–2437 (1990).
13. Beal, P. A. & Dervan, P. B. *J. Am. chem. Soc.* **114**, 4976 (1992).
14. Koh, J. S. & Dervan, P. B. *J. Am. chem. Soc.* **114**, 1470–1478 (1992).
15. Singleton, S. F. & Dervan, P. B. *J. Am. chem. Soc.* (in the press).
16. Pei, D., Corey, D. R. & Schultz, P. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9858–9862 (1990).
17. Koob, M., Grimes, E. & Szybalski, W. *Science* **241**, 1084–1086 (1988).
18. Strobel, S. A. & Dervan, P. B. *Nature* **350**, 172–174 (1991).
19. Povsic, T. J., Strobel, S. A. & Dervan, P. B. *J. Am. chem. Soc.* **114**, 5934–5941 (1992).
20. Ferrin, L. J. & Camerini-Otero, R. D. *Science* **254**, 1494–1497 (1991).

ADVERTISEMENTS

CLIX

Do you want to search the Cambridge Crystallographic Data Bank for Small Molecules capable of binding to known 3-D structures?

CLIX has the answer—quick & user friendly.

The algorithm was published in: *Proteins, Structures, Functions & Genetics* v 13:31–41.

CALL US TODAY!

CONTACT: Paul Davis or Mike Lawrence
CSIRO Division of Biomolecular Engineering,
343 Royal Parade
Parkville 3052
Victoria
AUSTRALIA
Tel: 613 342 4300 Fax: 613 342 4301
Email: paulc.rox.mcl.dbe.csiro.au

PACEMAKERSM

The multiple peptide synthesis and conjugation service

- High purity (up to 95%).
- Quantities for a range of uses.
- Single or multiple peptides.
- HPLC trace with each peptide.
- Terminal modification at no extra charge.
- Conjugation to carrier proteins.

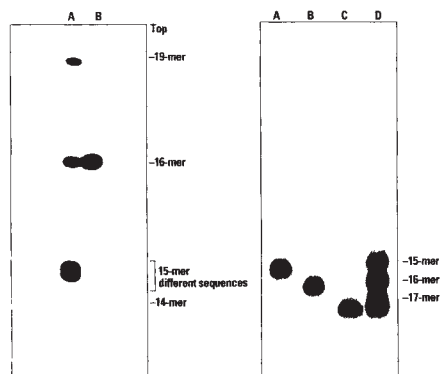
from £80 per peptide

AFFINITY Research Products Ltd.
GPT Business Park, Technology Drive,
Nottingham, NG9 2ND, UK.
Tel: +44/0 602 436100; Fax: +44/0 602 436300

Making ends meet

Featured this week — oligonucleotide purification and analysis kits, an instrument for automated oligo synthesis and a host of peptide-related products.

Two kits from United States Biochemical, SurePure and SureCheck, offer two new methods for either the **analysis or purification of oligonucleotide samples** (*Reader*



Be sure with USB's SurePure and SureCheck kits.

Service No. 101). In each kit, a deprotected oligonucleotide is applied to a specifically developed purification plate and eluted up the plate using the elution mixture supplied. The positions to which the oligonucleotides migrate are visualized using a UV light, with migration rate based on chain length. The \$150 (US) SureCheck kit allows rapid determination of the purity of an oligonucleotide sample, which is critical in polymerase chain reaction (PCR) and DNA sequencing. It includes all the reagents necessary for the analysis of 20 oligonucleotide samples. USB says that two samples can be analysed simultaneously on one plate so that a maximum of 40 samples can be checked by the reagents supplied in one kit. The \$185 (US) SurePure kit is intended for the rapid purification of oligonucleotides from failed sequences and contaminating salts, assuring the generation of target DNA by PCR and low background of DNA sequencing gels. The sample is applied to the bottom of the plate and following migration and visualization, the region of the plate containing the product oligonucleotide (the darkest band on the plate) is scraped away from the glass backing of the plate. The pure product is recovered by briefly soaking the powdered matrix in deionized water and filtering through a centrifuge filter. The kit includes all the reagents necessary to purify twelve 0.2- μ mol or six 1- μ mol scale syntheses.

The new **silver-based stain** from Integrated Separation Systems is intended for oligonucleotides such as PCR primers in addition to double-stranded DNA and RNA (*Reader Service No. 102*). The Oligo Staining System has a detection limit of 1–10 ng

per band and eliminates negative staining, which can be a common problem when using silver stain alone, ISS says. The stain can be used with Tris-glycine-SDS, TAE, TBE and TBE-urea gels.

■ Modes of synthesis

Primers and probes are available within 4 hours and DNA can be used without additional purification for most applications, including PCR, DNA sequencing and hybridization, says Beckman, with the introduction of its new **Oligo 1000 DNA synthesizer** (*Reader Service No. 103*). This instrument is designed to allow the user to produce oligonucleotides quickly, inexpensively and without special training. Beckman says that the Oligo 1000 uses 35 to 50 per cent less phosphoramidite. Other design features include the automatic determination of coupling efficiency (the user is notified should efficiency drop below an acceptable



Beckman's Oligo 1000.

level), the monitoring of reagent use (synthesis is stopped temporarily when an empty bottle is detected), the automatic monitoring of reagent age, flow rate and gas pressure, and an alarm system that alerts the user when a bottle is empty. With a cycle time of 4 min at 30-nmol synthesis scale, the Oligo 1000 delivers a 16 mer in about an hour, Beckman says.

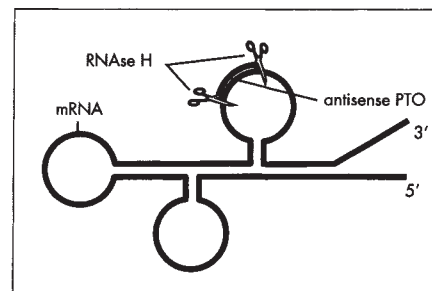
To meet your DNA synthesis needs, Pharmacia LKB Biotechnology has developed two new models of **DNA synthesizer** — the Gene Assembler Special and the Gene Assembler 4 Primers (*Reader Service No. 104*). The first of these instruments, the Gene Assembler Special, is designed and equipped to synthesize high-quality oligonucleotides labelled with biotin or fluorescein, phosphorylated or with 3'-modified nucleotides, phosphorothioated for antisense or RNA. In response to demand by users, Pharmacia is also releasing a set of new reagents specific for particular applica-

tions: Biodite, a biotin amidite for the labelling of primers and probes with biotin; AntiSense reagent for making phosphorothioated oligonucleotides for antisense research; and Phosphorylating amidite for 5' phosphorylations and the introduction of inosine at the 3' end of an oligonucleotide to avoid priming mismatches in PCR. The Gene Assembler 4 Primers is now equipped with four columns to ensure sufficient primer production capacity to support automated sequencing and PCR.

Advanced Biotechnologies is offering a complete range of **DNA synthesis-grade reagents** for the manual and automated synthesis of DNA (*Reader Service No. 105*). The line-up includes an activation reagent, deblocking agent, capping reagents A and B, an oxidation reagent and diluent. Each is rigorously quality-controlled to match the specific performance characteristics for the ABI synthesis, the company says. The reagents are suitable for use on the following models of ABI synthesizer: 391, 392, 394 and 380s fitted with large-bottle receptacles.

■ Making sense of antisense

In addition to its growing list of over 100 different pre-made and *in vivo* assayed antisense phosphorothioate oligonucleotides, Biometra now offers **custom antisense oligonucleotides** that are rigorously purified for *in vivo* applications (*Reader Service No. 106*). Protected custom antisense RNA oligonucleotides up to 50 residues in length are also available. Because the evidence shows RNAs to be more effective inhibitors than DNA due to their catalytic activity,



Antisense oligos to order.

Biometra says that they can be used in much smaller quantities. The company also offers a complete range of other synthesis services, including any type of coupling and modification of nucleic acids or peptides and synthesis of precursor molecules.

Riboblock allyl phosphoramidites, monomer building units in the synthesis of

ADVERTISEMENTS

CRYSTAL SCREEN™

Macromolecular crystallization reagent kit. Pre-formulated, ready to use. Proven the most efficient, rapid and effective tool for the crystallization of proteins and peptides. Also effective with oligonucleotides and small molecules.

For complete information on CRYSTAL SCREEN and all of our products for crystallization research please call or fax for a free brochure.

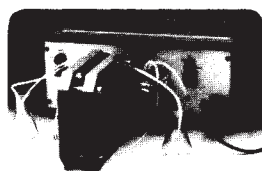
HAMPTON
RESEARCH

(800) 452-3899 Fax: (714) 789-8265
5225 Canyon Crest Dr. Suite 71-336
Riverside, CA 92507 U.S.A.

Reader Service No.22

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.21

DNA FINGERPRINTING

Fresenius AG, Bad Homburg is offering the following oligonucleotides **exclusively**:



sirf 41 "Sherlock Holmes"

sirf 42 "Dr. Watson"

sirf 11 "Miss Marple"

Multi-Locus-Probes

(GTG)₅ · (GATA)₄ · (GACA)₄

(GGAT)₄ · (TCC)₅ · (CA)₈

Fresenius
DIA•GNOSTICS

6370 Oberursel, F.R.G. · Telephone: + 6171/60 2443
Telefax: + 6171/60 2317 · Postbox 1809

Reader Service No. 9

nuclease-resistant oligoribonucleotides, are now available from Boehringer Mannheim and should be of interest to researchers involved with antisense technology (*Reader Service No. 107*). The company says that the allyl moiety at the 2' hydroxyl group of the ribose residues renders Riboblock polymers completely resistant to DNA- and RNA-specific nucleases — a critical factor with regard to the biological application of any antisense oligo. Allyl-modified RNAs in comparison to methyl-modified analogues also exhibit a reduced binding capacity to non-desired proteins. They are suitable for automated solid-phase synthesis using the phosphite-triester method.

Immunology update

IBL Research Products, a wholly owned subsidiary of International Biotechnology Laboratories, announces the availability of recombinant Epstein-Barr Virus (EBV) viral capsid antigen peptides for the detection of EBV antibodies (*Reader Service No. 108*).



Recombinant EBV antigens.

These peptides are purified to greater than 90 per cent. IBL RPC's recombinant antigens, L1, L2 and L1-MA, have been developed in-house by the parent company using proprietary technology.

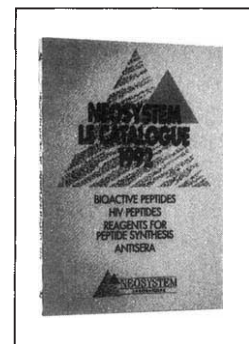
For researchers who need **custom polyclonal anti-peptide antibodies** in a hurry, Multiple Peptide Systems says it can help (*Reader Service No. 109*). In addition to the standard monoclonal and polyclonal anti-peptide antibody services, new proprietary immunization and bleed techniques allow MPS to offer polyclonal anti-peptide antibodies 33 per cent faster than the standard service. The service can be performed on your purified peptide or, MPS says, it

can synthesize the peptide (synthesis includes mass spectral analysis and HPLC purity checks). The full service includes: 20 mg of 90–95% pure peptide, approximately 150 ml of ELISA-tested crude rabbit serum from three rabbits, 5 ml of affinity-purified serum and the column used for affinity purification with the peptide attached.

Peptide profile

Affiniti Research Products has launched a new **PACemaker custom peptide synthesis and conjugation service** that utilizes advances in multi-format, solid-phase modes of synthesis and state-of-the-art purification procedures (*Reader Service No. 110*). Three grades of peptide purity are offered: ≥70% for general immunological applications, ≥80 % for initial screening and bioassays and ≥95% for stringent pharmacological applications and structural studies. Each peptide is analysed by reversed-phase HPLC and a chromatogram accompanies the lyophilized product. Standard quantities are 1 mg and 5 mg and peptides up to 20 amino acid residues in length may be synthesized using Affiniti's routine service. Larger quantities are available on request. Peptide modification, including N-terminal acetylation or C-terminal amidation, is provided at no extra cost, the company says.

Neosystem Laboratoire boasts that it has all the **peptides** you need! The latest edition of the company's catalogue includes hundreds of references for neuropeptides, brain/gut peptides, specific protein kinase C inhibitors, tachykinins and other bioactive peptides, peptides relevant to AIDS research, human immunodeficiency virus, simian immunodeficiency virus and human T-lymphotrophic virus epitopes (*Reader Service No. 111*). Reagents for peptide synthesis, antisera for radio-immunoassay are also featured.



Assorted peptides.

The Swiss Bachem group announces the availability of its new product **catalogue** (*Reader Service No. 112*). It should be useful as a source of amino acid derivatives, resins, enzyme substrates and inhibitors, cytokines, growth factors, neurotoxins and other bioactive peptides. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. For more information, fill in the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

Professional Custom Services

DNA SYNTHESIS
PROTEIN SEQUENCING
PEPTIDE SYNTHESIS

Rapid Service • Guaranteed Quality • Ample Quantity
Competitive Prices • Free Quotations
World-wide Delivery

V.G. Services

1200 Wonderland Rd. S. Bldg. #7 Unit #2
London, Ontario Canada N6L 1A8

CALL/FAX: (519) 652-3758 ALT. FAX: (519) 652 0053

Reader Service No.24